STIC-Biotech/ChemLib

99802

09/729,264 11/26/00 Welcher et al.

search seq id nos 1, 3, and 5 against us patent and us patent application databases.

search seq id nos 2, 4, and 6 against us patent and us patent application databases.
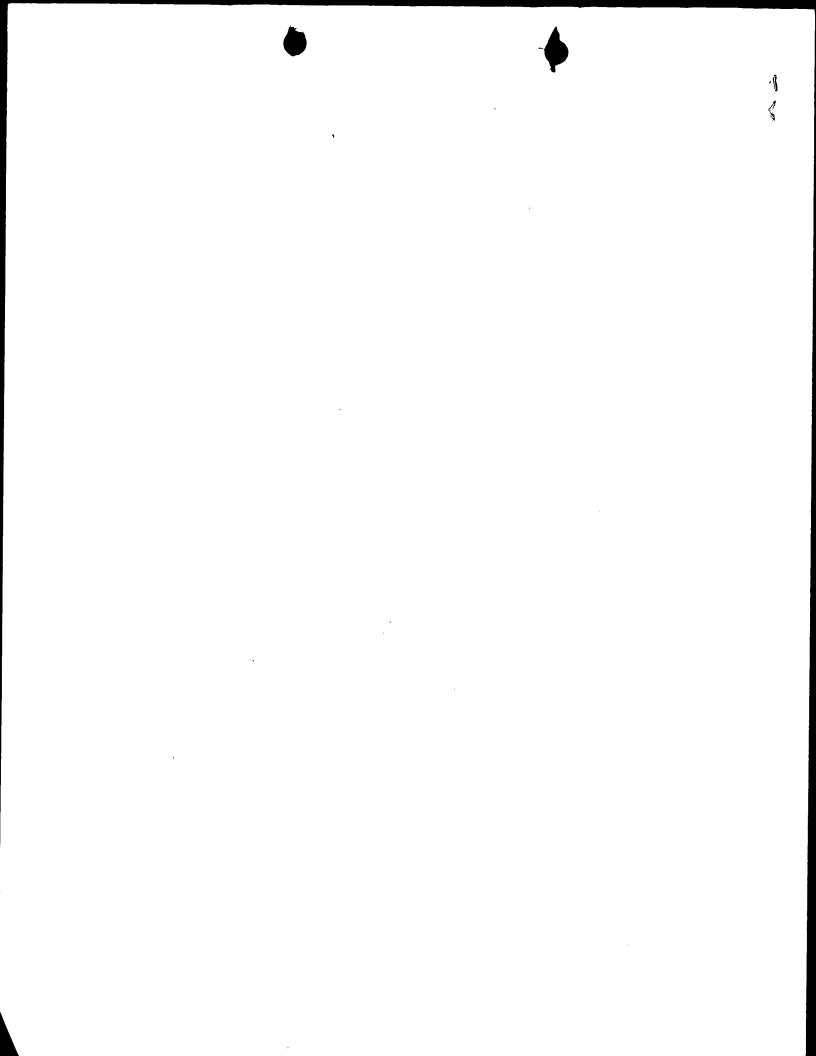
Thanks,
Brian Whiteman, 11e12
Patent Examiner - Art Unit 1635
United States Patent and Trademark Office
Crystal Mall 1, 11A16
(703) 305-0775

Mary Jane Ruhl
Tech. Info. Specialist, STIC
TC-1600
CM-1, Room 6A-06
Phone: 605-1155

Searcher:_____
Phone:_____
Location:_____
Date Picked Up:_____
Date Completed: 4/30/03
Searcher Prep/Review:_____
Clerical:_____
Online time:_____

TYPE OF SEARCH:
NA Sequences:_____
AA Sequences:_____
Structures:_____
Bibliographic:_____
Litigation:_____
Full text:_____
Patent Family:_____
Other:_____

VENDOR/COST (where applic.)
STN:_____
DIALOG:_____
Questel/Orbit:_____
DRLink:_____
Lexis/Nexis:_____
Sequence Sys.:_____
WWW/Internet:_____
Other (specify):_____

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:24:32 ; Search time 36.7435 seconds
(without alignments)
2142.147 Million cell updates/sec

Title: US-09-729-264-2
Perfect score: 2059
Sequence: 1 MGLVIFLHGSGSGNEVIEGP.........HPQASFNLASPEKVSNTTVV 382

Scoring table: BLOSUM62
  Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Post-processing: Minimum Match 0%
  Maximum Match 100%
  Listing first 45 summaries

Database : SPTREMBL_21:*
 1: sp_archea:*
 2: sp_bacteria:*
 3: sp_fungi:*
 4: sp_human:*
 5: sp_invertebrate:*
 6: sp_mammal:*
 7: sp_mhc:*
 8: sp_organelle:*
 9: sp_phage:*
 10: sp_plant:*
 11: sp_rodent:*
 12: sp_virus:*
 13: sp_vertebrate:*
 14: sp_unclassified:*
 15: sp_rvirus:*
 16: sp_bacteriap:*
 17: sp_archeap:*

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1694	82.3	315	4	Q9NSI5	Q9ns15 homo sapien
2	575.5	28.0	270	11	Q9D8G2	Q9d8g2 mus musculu
3	141	6.8	168	11	Q9D9I2	Q9d9i2 mus musculu
4	136.5	6.6	173	11	Q9D4K2	Q9d4k2 mus musculu
5	135	6.5	1496	4	Q92626	Q92626 homo sapien
6	133	6.5	1427	13	Q91562	Q91562 xenopus lae
7	132.5	6.4	344	13	Q9DF61	Q9df61 gallus gall
8	131	6.4	697	11	Q9DBP0	Q9dbp0 mus musculu
9	131	6.4	697	11	Q9Z290	Q9z290 mus musculu
10	129.5	6.3	164	13	Q22048	Q22048 caenorhabdi
11	127.5	6.2	315	13	Q9DGI5	Q9dgi5 gallus gall
12	127.5	6.2	344	13	Q99PJ0	Q99pj0 gallus gall
13	127.5	6.2	344	13	O93242	O93242 gallus gall
14	127	6.2	871	11	O61987	O61987 mus musculu
15	127.5	6.2	881	11	O61988	O61988 mus musculu
16	127	6.2	197	5	O17641	O17641 caenorhabdi
17	127	6.2	1445	11	Q63155	Q63155 rattus norv
18	126.5	6.1	869	4	O15146	O15146 homo sapien
19	126	6.1	313	13	O57596	O57596 gallus gall
20	125.5	6.1	166	5	Q95QY1	Q95qy1 caenorhabdi
21	125.5	6.1	6632	5	O01761	O01761 caenorhabdi
22	123.5	6.0	344	4	Q9P121	Q9p121 homo sapien
23	123.5	6.0	188	4	Q96AA2	Q96aa2 homo sapien
24	122.5	5.9	6620	5	O18238	O18238 caenorhabdi
25	122.5	5.9	455	4	Q9UIR0	Q9uir0 homo sapien
26	120	5.8	959	4	Q9N9Y9	Q9n9y9 drosophila
27	120	5.8	968	5	Q9W4T9	Q9w4t9 drosophila
28	119	5.8	975	5	O97174	O97174 rattus norv
29	119	5.8	975	11	Q9JK39	Q9jk39 mus musculu
30	119	5.7	275	11	Q9H1X9	Q9h1x9 homo sapien
31	117.5	5.7	390	4	Q96RW7	Q96rw7 homo sapien
32	117.5	5.7	5636	5	O02838	O02838 sus scrofa
33	117	5.7	325	5	O02838	O02838 drosophila
34	116.5	5.7	345	5	Q9W4U1	Q9w4u1 drosophila
35	116	5.6	2221	5	Q9U1M1	Q9u1m1 drosophila
36	116	5.6	4370	4	Q9H3V5	Q9h3v5 homo sapien
37	116	5.6	6632	5	O17362	O17362 caenorhabdi
38	115.5	5.6	196	5	Q9H0T1	Q9h0t1 homo sapien
39	115.5	5.6	152	5	Q9XVX3	Q9xvx3 caenorhabdi
40	114	5.6	46	13	Q16861	Q16861 homo sapien
41	114	5.5	327	4	Q96IO7	Q96io7 homo sapien
42	114	5.5	569	11	Q920G3	Q920g3 mus musculu
43	114	5.5	1395	5	O44924	O44924 drosophila
44	114	5.5	1395	5	Q9W213	Q9w213 drosophila
45	114	5.5	2673	4	Q96SC3	Q96sc3 homo sapien

ALIGNMENTS

RESULT 1

Q9NSI5

ID Q9NSI5 PRELIMINARY; PRT; 315 AA.
AC Q9NSI5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IGSF5 protein (Fragment).
GN IGSF5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.S.,
RA Toyoda A., Ishii K., Totoki Y., Choi D.K., Soeda E., Ohki M.,
RA Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Yaspo M.L.,
RA Lehrach H., Reinhardt R., Nizetic D., Francis F.,
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL163280; CAB90047.1; -.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR0003006; Ig_MHC.
DR Pfam; PF00047; Ig; 2.
DR SMART; SM00409; IG; 2.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
FT NON_TER 1 1
FT NON_TER 315 315
SQ SEQUENCE 315 AA; 34599 MW; B69DCB23570FA3D1 CRC64;

```
Query Match
Best Local Similarity   82.3%;  Score 1694;  DB 4;  Length 315;
Matches  314;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    10 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 69
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 60

Qy    70 SQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVVA 129
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 SQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVVA 120

Qy   130 ENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALFPQSNGTLT 189
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALFPQSNGTLT 180

Qy   190 CVATWKSLKARKSATVNLIVIRCPQDTTGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGT 249
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CVATWKSLKARKSATVNLIVIRCPQDTTGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGT 240

Qy   250 MLLTPTCTLTLIRCCCCRRKRRCCGCNCCCRCCFCCRRKRRGFRIQFQKKSEKEKTNKETETES 309
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 MLLTPTCTLTLIRCCCCRRKRRCCGCNCCCRCCFCCRRKRRGFRIQFQKKSEKEKTNKETETES 300

Qy   310 GNENSGYNSDEQKTT 324
            ||||||||||||||
Db   301 GNENSGYNSDEQKTT 315
```

RESULT 2
Q9D8G2

```
ID   Q9D8G2;      PRELIMINARY;      PRT;   270 AA.
AC   Q9D8G2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2010003D20Rik protein (RIKEN cDNA 2010003D20 gene).
GN   2010003D20Rik.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK008060; BAB25436.1; -.
DR   EMBL; BC004806; AAH04806.1; -.
DR   MGD; MGI:1919308; 2010003D20Rik.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   SMART; SM00409; IG; 1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
SQ   SEQUENCE   270 AA;  29604 MW;  A39C273DAA950DFE CRC64;
```

```
Query Match
Best Local Similarity   28.0%;  Score 575.5;  DB 11;  Length 270;  Pred. No. 2.3e-44;
Matches   35.1%;  Conservative   42;  Mismatches   73;  Indels   6;  Gaps   6;

Qy     3 LVIF--LHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPME 60
Db    12 LVILAQLTASGSSYQIIEGPQNVTVLRDSEAHFNCTVTHGWKLLMWTLNQMVVLSLTTGG 71

Qy    61 PIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELF 120
Db    72 PIITNNRRTYASYNSTDSFISELIIHDVQPSDSGSVQCSLQNSHGFGSAFLSVQ------ 125

Qy   121 IPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILAL 180
Db   126 ------------------------------------------------------------ 125

Qy   181 TPQSNGTLTCVATWKSLKARKSATVNLIVIRCPQDTTGGINIPGVLSSLPSLGFSLPTWG 240
Db   126 ------------------------------------------------------------ 125

Qy   241 KVGLGLAGTIMLLPTCTLTLIRCCCCRRKRRCCGCNCCCRCCFCCRRKRRGFRIQFQ---KKSE 297
Db   140 IILLAVAFSLLLLIIIIVLIIIFCCC--------------------CASRREKEESTYQNEIRKSA 184

Qy   298 KEKTNK---ETETESGNENSGYNSDEQKTTDTASLPPKSCESSDPEQRNSSCCGPHORAD 354
Db   185 NMRTNKADPETRLKSCKENYGYSSDEAKAAQTASLPPKSAEVSLPEKRSSSL--PYQELN 242

Qy   355 QRPPRPASHPQOASFNLASPEKVSNTTVV 382
Db   243 KHQPGPATHPRVSFDIASPQKVRNVTLV 270
```

RESULT 3
Q9D9I2

```
ID   Q9D9I2;      PRELIMINARY;      PRT;   168 AA.
AC   Q9D9I2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   4931420D14Rik protein.
GN   4931420D14Rik.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
```

STRAIN=C57BL/6J; TISSUE=TESTIS;

RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK006892; BAB24782.1; -.
DR MGD; MGI:1913992; 4931420D14Rik.
SQ SEQUENCE 168 AA; 18931 MW; 7A2BD279612A5E94 CRC64;

Query Match   6.8%; Score 141; DB 11; Length 168;
Best Local Similarity 32.0%; Pred. No. 5.3e-05;
Matches 39; Conservative 20; Mismatches 49; Indels 14; Gaps 6;

Qy 256 CTLTIRCCCRRRCCGCNC--CCRCCFCCRRKRGFR----IQFQKKSEKETNKETETESG 310
         ||||| | |||||| ||||||| |||||| ||||||| |||:||
Db 112 CSLSRSCCCCRCRCC-CHCRCRC--CCSRSRFRSRTTLRKFFQITEKGEQSLQRRIRRQ 111

Qy 311 NENSGYNSDEQKTDTASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASFNL 370
         |::|:| || :|: |::| ||:||:   :|
Db 112 LTRSQLELIEPE--PTMALEPSEITVAFFSHKNANVSDP----EEVPPCLDSDPFPNGDL 165

Qy 371 AS 372
         ||
Db 166 AS 167

RESULT 4
ID Q9D4K2
AC Q9D4K2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 4931420D14Rik protein.
GN 4931420D14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki H., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK016467; BAB30253.1; -.
DR MGD; MGI:1913992; 4931420D14Rik.
SQ SEQUENCE 173 AA; 19581 MW; 147B6F155AC29FDF CRC64;

Query Match   6.6%; Score 136.5; DB 11; Length 173;
Best Local Similarity 31.5%; Pred. No. 0.00014;
Matches 40; Conservative 16; Mismatches 52; Indels 19; Gaps 6;

Qy 256 CTLTIRCCCRRRCCGCNC--CCRCCFCCRRKRGFRIQFQKKSEKEKTNKETETESGNENS 314
         ||||| | |||||| ||||||| |||||| || || |||:||
Db 55 CSLSRSCCCCRCRCC-CYCRCRC--CCSRSRFRSRTTLRVRDPKFFQ--ITEKGEQSL 109

Qy 315 GYNSDEQKTD------------TASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQ 365
          ||:|:| |::| ||:||:   :|
Db 110 QRRIRQLTRSQLELIEPEPTMALEPSEITVAFFSHKNANVSDP----EEVPPCLDSDPF 165

Qy 366 ASFNLAS 372
        ::||||
Db 166 PNGDLAS 172

RESULT 5
ID Q92626
AC Q92626;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE KIAA0230.
GN KIAA0230.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BONE MARROW;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain.";
RL DNA Res. 3:321-329(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RX MEDLINE=95048383; PubMed=7959781;
RA Weiler S.R., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,
RA Trent J.M.;
RT "Assignment of a human melanoma associated gene MG50 (D2S448) to
RT chromosome 2p25.3 by fluorescence in situ hybridization.";
RL Genomics 22:243-244(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=MELANOMA;
RA Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;
RT "Identification of a novel melanoma gene (MG50) - likely the gene for
RT IL-1 receptor antagonist - which encodes epitopes recognized by human
RT cytolytic T lymphocytes.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; D86983; BAA13219.1; -.
DR EMBL; AF200348; AAF06354.1; -.
HSSP; P05164; 1CXP.
DR InterPro; IPR002007; Anim_peroxidase.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR000483; LRR_Cterm.
DR InterPro; IPR000372; LRR_Nterm.
DR InterPro; IPR003591; LRR_typ.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF03098; An_peroxidase; 1.
DR Pfam; PF00047; Ig; 4.
DR Pfam; PF00560; LRR; 5.
DR Pfam; PF01463; LRRCT; 1.
DR Pfam; PF00093; vwc; 1.
DR PRINTS; PR00457; ANPEROXIDASE.
DR SMART; SM00408; IGc2; 4.
DR SMART; SM00082; LRRCT; 1.
DR SMART; SM00013; LRRNT; 1.
DR SMART; SM00369; LRR_TYP; 4.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC; UNKNOWN_1.
KW Immunoglobulin domain.

FT   NON_TER
SQ   SEQUENCE   1496 AA;   167209 MW;   E9B9A7069BF1ABFF CRC64;
Query Match
Best Local Similarity   25.0%;   Score 135;   DB 4;   Length 1496;
Matches   73;   Conservative   36;   Mismatches 127;   Indels   56;   Gaps 15;

```
Qy     9 GSGSGNEVIEGPQNATVLKGSQARNCTVS-QGWKLIMWALSDMVLSVRPMEPIITNDR 67
           :: ::     ::   :::: : :: :::    :::::::::::::::::::::::::
Db   354 GSPARPFVIQPQNTEVLVGGESVTLECSATGHPPPRISWTRGDRTLPFVPDPRVNITPS-- 411

Qy    68 FTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNS--RLHGSAYLITVQVMGELFIPSVN 125
Db   412 -------GG----LYIQNVVQGDSGEYACSAATNNIDSVHATAFIIVQALPQFTVTPQD 458

Qy   126 LVVAAENEPCEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQS 184
Db   459 RVVIEGQTVDFQC-EAKGNPPPVIAWTKGGSQLSVDRHHLVLSSGTLR--ISGVALHDQ- 514

Qy   185 NGTLTCVATWKSLKARKSATVNLTV----------IRCPQDT----GGGINIPGVLSSLPSL 232
Db   515 -GQYECQAV--NIIGSQKVAHLTVQPRVTPVFASIPSDTTVEVGANVQLP-----CSSQ 566

Qy   233 GFSLP--TWGKVLGL---AGTMLLTPTCTLTI---------------RCCCCRRRCCG 271
Db   567 GEPEPAITWNKDGVQVTESGKPHISPEGFLFINDVGPADAGRYECVARNTIG 618
```

RESULT 6

ID   Q91562
AC   Q91562;
DT   01-NOV-1996 (TREMBLrel. 01, Created)
DT   01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE   Tumor suppressor.
GN   XDCCA.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OX   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95113183; PubMed=7813784;
RA   Pierceall W.E., Reale M.A., Candia A.F., Wright C.V., Cho K.R.,
RA   Fearon E.R.;
RT   "Expression of a homologue of the deleted in colorectal cancer (DCC)
RT   gene in the nervous system of developing Xenopus embryos.";
RL   Dev. Biol. 166:654-665(1994).
DR   EMBL; U10986; AAA70168.1;
DR   HSSP; P40189; 1BQU.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00408; IG2; 3.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE   1427 AA;   156533 MW;   61FEA12C8A674972 CRC64;
PRELIMINARY;   PRT;   1427 AA.

Query Match
Best Local Similarity   25.7%;   Score 133;   DB 13;   Length 1427;
Matches   59;   Conservative   33;   Mismatches   90;   Indels   48;   Gaps 13;

```
Qy    17 IEGPQNATVLKGSQARNCTVSQGW-KLIMWALSDMVVLSVRPMEPIITNDRFTSQRYD 74
Db   243 LQRPSNVVAIEGQDAVLECAVS-GYPTPTIVWMQGD------EPVPIRTR---KYS 288
```

RESULT 7

ID   Q9DF61
AC   Q9DF61;
DT   01-MAR-2001 (TREMBLrel. 16, Created)
DT   01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE   OBCAM alpha 1 isoform.
GN   OBCAM.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20499204; PubMed=11042360;
RA   Lodge A.P., Howard M.R., McNamee C.J., Moss D.J.;
RT   "Co-localisation, heterophilic interactions and regulated expression
RT   of IgLON family proteins in the chick nervous system.";
RL   Brain Res. Mol. Brain Res. 82:84-94(2000).
DR   EMBL; AF292934; AAG01877.1;
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; Ig; 3.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IG2; 3.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain.
SQ   SEQUENCE   344 AA;   37531 MW;   37FE6051CBF0E7B4 CRC64;
PRELIMINARY;   PRT;   344 AA.

Query Match
Best Local Similarity   25.5%;   Score 132.5;   DB 13;   Length 344;
Matches   70;   Conservative   38;   Mismatches 104;   Indels   63;   Gaps 16;

```
Qy     4 VIFLHGSG-----SGNEVI-EGPQNATVLKGSQARNCTVSQGWKLIMWALSDMVVL----- 54
Db    21 LLFLVPAGVPVRSGDATFPKAMDNVTVRQGESATLRCTVDDRVRRVAV-LNRSTILVAGN 79

Qy    55 ---SVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRL 105
Db    80 DKWSIDNRVVILSN--TKTQY------SIKIHNVDYDEGPYTCSVQTDNHPKTSRV 128

Qy   106 HGSAYLITVQVMGELFIPSVNLVVAAENEPCEVTCL----PSHWTRLPDISWEIGLLVSHSS 161
Db   129 H----LIVQVPPQIVNISSDITVNEGSSVTLMCLAFGRPE----PTVTWR--HLSGKG 176

Qy   162 YFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-----IRCPQDT 216
Db   177 QGFVSEDEYLE----IIGITREQSGEYCSAV-NDVAVPDVRKVKVTVNPPYISNAKNT 231

Qy   217 GGGINIPGVL-----SSLPSLGFSLPTWGKVGLGLA 247
Db   232 GASVGQKGILQCEASAVPVAEFQ--WFKEDTRIA 263
```

```
Qy    75 QGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSAYLITVQVMGELFIPSVNLVVAAENE 132
Db   289 VLGG--SNLLISNVTDDDAGAYTCVATYKNENTSFSADLIVMVPPQFLNHPANLYAYESM 346

Qy   133 PCEVTCL----PSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGT 187
Db   347 DIEFECAVSGKPS--------PIVKWIKNGEVVIPSDYFQIVDGSNLR----ILGLVKSDEGY 397

Qy   188 LTCVA--------TWKSLKARKSATVNLTVI-RCPQDIGGGINIPGVLSS 228
Db   398 YQCIAENEAGNIQTYAQLIIPDPAVPSSSILPSAPRDV----VPVLVSS 442
```

RESULT 8

ID	Q9DBP0	PRELIMINARY;	PRT;	697 AA.

AC   Q9DBP0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Solute carrier family 34 (sodium phosphate), member 2.
GN   SLC34A2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LUNG;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shibagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK004832; BAB23600.1; -.
DR   MGD; MGI:1342284; Slc34a2.
DR   Pfam; PF02690; Na_Pi_cotrans; 1.
DR   InterPro; IPR003841; Na_Pi_cotrans.
DR   TIGRFAMS; TIGR01013; 2a58; 1.
SQ   SEQUENCE   697 AA; 76244 MW; 2A7B9384857EF16F CRC64;

Query Match   6.4%; Score 131; DB 11; Length 697;
Best Local Similarity   20.6%; Pred. No. 0.0028;
Matches   89; Conservative   51; Mismatches 123; Indels 168; Gaps 23;

```
Qy       3 LVIFLHGSGSGNEVIEGPQNATVLK--------------------------GSQARFNCTVSQG  40
             ::|: | |||                                      |: :|||:
Db     382 MIVKLLGS-----VLRG-QVATVIKKTLNTDFPFPFAWLTGYLAILVGAGMTFIVQSSSV  435

Qy      41 WKLLIMWALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHHNVEPSDSGNIRCSL  100
           |||:||  ||     |:| |||| | :| | | | :| | : |  |  | ||  : |
Db     436 FTSAMTPLIGIGVISIERAYPLTLGSNI------GTTTTAIALASPGNI--LRSSL  485

Qy     101 QNSRLHGSAYLTVQVMGELFIPSVNLVVAAENEPCEVTCLPSHWTRLP-------DIS--  150
           | |  | |  :| | |                          : ||| |      | :
Db     486 QIALCH--------------------------------------FFFNISGI-----  524

Qy     151 --WELGLLIVSHSSYFVFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLT  208
             ::|  : |  | : | :   |||   |   |   ||  :|:|  |   |: ||  ||
Db     525 YRW-----FAVFYLIFFFVTP-----LIVFGLS--LAGWPVLVGVGVPIILLL  566

Qy     209 VI-------RCPQDTGGGINIPGVL----SSLPSLGFSLPTWGKVGLGLAGTIMLLTPTC  256
           |:       :|  |||||:||  |:    :  | | | | | | ||: || |  || |
Db     567 LIVLCLRMLQFRCPR-----ILPLKIRDWNFLPLWMHSLKPWDNV------ISLATTC  613

Qy     257 TLTIRCCCCRRRCCG-----CNC-CCRCCFCCRRKRGFRIQFQKKSEKKTNKETETESG  310
           |||:||   |  |||     |:| |||||  |   |  |:|:|                :
Db     614 -FQRCCCCCRVCCMVCGCKCCRCSKCCR--------------------------DQG  648

Qy     311 NENSGYNSDEQKTTDTIASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASFNL  370
           :   | | |   : |  ::  |  | |: |||  : | | |  | | :  |    |  |
Db     648 LSMKALSNTTV  —
```

RESULT 9

ID	Q9Z290	PRELIMINARY;	PRT;	697 AA.

AC   Q9Z290;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Type IIb Na/phosphate cotransporter.
GN   SLC34A2 OR NPT2B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NMRI; TISSUE=SMALL INTESTINE;
RX   MEDLINE=99045724; PubMed=9826740;
RA   Hilfiker H., Hattenhauer O., Traebert M., Forster I., Murer H.,
RA   Biber J.;
RT   "Characterization of a murine type II sodium-phosphate cotransporter
RT   expressed in mammalian small intestine.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:14564-14569(1998).
DR   EMBL; AF081499; AAC80007.1; -.
DR   MGD; MGI:1342284; Slc34a2.
DR   InterPro; IPR003841; Na_Pi_cotrans.
DR   Pfam; PF02690; Na_Pi_cotrans; 1.
DR   TIGRFAMS; TIGR01013; 2a58; 1.
SQ   SEQUENCE   697 AA; 76286 MW; 839E5CCB0F565265 CRC64;

Query Match   6.4%; Score 131; DB 11; Length 697;
Best Local Similarity   20.6%; Pred. No. 0.0028;
Matches   89; Conservative   51; Mismatches 123; Indels 168; Gaps 23;

```
Qy       3 LVIFLHGSGSGNEVIEGPQNATVLK--------------------------GSQARFNCTVSQG  40
             ::|: | |||                                      |: :|||:
Db     382 MIVKLLGS-----VLRG-QVATVIKKTLNTDFPFPFAWLTGYLAILVGAGMTFIVQSSSV  435

Qy      41 WKLLIMWALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHHNVEPSDSGNIRCSL  100
           |||:||  ||     |:| |||| | :| | | | :| | : |  |  | ||  : |
Db     436 FTSAMTPLIGIGVISIERAYPLTLGSNI------GTTTTAIALASPGNI--LRSSL  485

Qy     101 QNSRLHGSAYLTVQVMGELFIPSVNLVVAAENEPCEVTCLPSHWTRLP-------DIS--  150
           | |  | |  :| | |                          : ||| |      | :
Db     486 QIALCH--------------------------------------FFFNISGI-----  524

Qy     151 --WELGLLIVSHSSYFVFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLT  208
             ::|  : |  | : | :   |||   |   |   ||  :|:|  |   |: ||  ||
Db     525 YRW-----FAVFYLIFFFVTP-----LIVFGLS--LAGWPVLVGVGVPIILLL  566

Qy     209 VI-------RCPQDTGGGINIPGVL----SSLPSLGFSLPTWGKVGLGLAGTIMLLTPTC  256
           |:       :|  |||||:||  |:    :  | | | | | | ||: || |  || |
Db     567 LIVLCLRMLQFRCPR-----ILPLKIRDWNFLPLWMHSLKPWDNV------ISLATTC  613

Qy     257 TLTIRCCCCRRRCCG-----CNC-CCRCCFCCRRKRGFRIQFQKKSEKKTNKETETESG  310
           |||:||   |  |||     |:| |||||  |   |  |:|:|                :
Db     614 -FQRCCCCCRVCCMVCGCKCCRCSKCCR--------------------------DQG  648

Qy     311 NENSGYNSDEQKTTDTIASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASFNL  370
           :   | | |   : |  ::  |  | |: |||  : | | |  | | :  |    |  |
Db     648 EE------EEEKEQD--IPVKASGAFDNAAMSKEC------------------QDEGKGQVEV  685

Qy     371 ASPEKVSNTTV  381
           | ||:|||||
Db     686 LSMKALSNTTV  696
```

```
RESULT 10
ID   Q22048           PRELIMINARY;      PRT;   164 AA.
AC   Q22048;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   T01B7.8 protein.
GN   T01B7.8.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sims M.A.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z66499; CAA91301.1; -.
DR   InterPro; IPR001450; 4Fe4s_ferredoxin.
DR   InterPro; IPR001271; Defensin_mammal.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001007; VWF_C.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; UNKNOWN_1.
DR   PROSITE; PS00269; DEFENSIN; UNKNOWN_1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01208; VWFC; UNKNOWN_1.
SQ   SEQUENCE   164 AA;  16499 MW;  C002D48D36C9FCED CRC64;

Query Match             6.3%;  Score 129.5;  DB 5;  Length 164;
Best Local Similarity   32.1%;  Pred. No. 0.00058;
Matches   43;  Conservative   11;  Mismatches   53;  Indels   27;  Gaps   6;

Qy    153 LGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRC 212
Db      6 LAILLAIGTFIAV----SQVQSAV-----LPVSSTELATVGTDVSTASTAIDTLGNSSSRV 57

Qy    213 PQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLLAGTMLITPICTLTIRCCCRRRCCGC 272
Db     58 KRQGGCGCCCGCG--------------GCCGCGGGGGG-ᒻCGCCCCRPRCCCCCRRCCTC 101

Qy    273 --NCCC-RCCFPCCR 283
Db    102 CRTCCCTRCCTCCR 115


RESULT 11
ID   Q9DGI5           PRELIMINARY;      PRT;   315 AA.
AC   Q9DGI5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CEPU-Se alpha 1 isoform.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Lodge A.P.; McNamee C.J.; Howard M.R.; Reed J.E.; Moss D.J.;
RT   "Characterisation of CEPU-Se, a secreted isoform of the IgLON family
RT   protein CEPU-1.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF292936; AAG01879.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain.
SQ   SEQUENCE   315 AA;  34606 MW;  68C5D27F0DDC6FB2 CRC64;

Query Match             6.2%;  Score 127.5;  DB 13;  Length 315;
Best Local Similarity   25.4%;  Pred. No. 0.0021;
Matches   68;  Conservative   41;  Mismatches   98;  Indels   61;  Gaps   15;

Qy      4 VIFLHGSG----SGNEVI-EGPQNAIVLKGSQOARFNCTVSQ------GW--KLIMWALSD 50
Db     21 LLFLPAGVPVPVRSGDAITFPKAMDNVTVRQGESATLRCSVDNRVTRVAWLNRSSILYAGND 80

Qy     51 MVVLSVRPMEPIIINDRFTSQQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSR 104
Db     81 KWCLDPRVVLLANTKTQYSIQ-----------IHDVDYDEGPYTCSVQTDNHPKTSR 127

Qy    105 LHGSAYLTVQVMGELFIPSVNLVAVAENEPCEVTLCLPSHWTRLPD--ISNELGLLVSHSSY 162
Db    128 VH----LIVQVSPKITEISSDISINEGGNVSLTCIA---TGRPDPTITWR--HISPRAV 177

Qy    163 YFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-------IRCPQDTG 217
Db    178 GFISEDEYLE----ITGITREQSGEYECSAS-NDVAAPVVQRVKVTVNYPPYISDAKSTG 232

Qy    218 GGINIPGVL---SSLPSLGFSLPTWGK 241
Db    233 VPVGQKGILMCEASAVPSADFQ---WYK 257


RESULT 12
ID   Q99PJ0           PRELIMINARY;      PRT;   344 AA.
AC   Q99PJ0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Neurotrimin.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR; TISSUE=BRAIN;
RA   Kim T.H.; Choi S.C.; Kim J.; Jeon J.W.; Kim K.D.; Lee S.H.;
RT   "Cloning and expression of mouse neurotrimin gene in the developing
RT   nervous system.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF282980; AAK00276.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain.
SQ   SEQUENCE   344 AA;  37924 MW;  3ECC6D5E86C5C17D CRC64;

Query Match             6.2%;  Score 127.5;  DB 11;  Length 344;
Best Local Similarity   25.3%;  Pred. No. 0.0023;
Matches   66;  Conservative   39;  Mismatches   109;  Indels   47;  Gaps   14;
```

Left column

```
Qy    4 VIFLHGSG----SGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRP 58
         : ||: ||      | | :|  :||:||  |  | || :| | || : ||  :|  |
Db   21 LILFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDNRVTRVAWLNRSTI----- 74

Qy   59 MEPIITNDRF-TSQRYDQGGNFTSE--MIIHNVEPSDSGNIRCSLQ------NSRLHGSA 109
          | ||       |||  ||  | |:     ||| ||||||||||||        ||:||||
Db   75 -PYAGNDKWCLDPRVVLLGNTQTQYSIEIQNVDVYDEGPYTCSVQIDNHPKTSRVH---- 129

Qy  110 YLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYFVPEPS 169
          :||||| |||||||||  :    :  | |  ||||||| : ||: | || ||:||||
Db  130 -LIVQVSPKIVELISSDISINEGNNISLTCIATGRPE-PTVTWR---HISPKAVGEVSEDE 184

Qy  170 DLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP----QDTGGGINIPG 224
          :|:|||| |  | | | | ||| | | | | |  |  : | | |     | |||||| |
Db  185 YLQ----IQGITRFQSGEYECSAS-NDVAAPVVPRVKTVNYPYFSEAKGTGVPVGQKG 239

Qy  225 VL----SSLPSLGFSLPTWGK 241
          :|       ||:||||||||
Db  240 TLQCEASAVPSAEFQ---WFK 257
```

RESULT 13

```
ID   O93242        PRELIMINARY;      PRT;   344 AA.
AC   O93242;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CEPU-1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kimura Y., Shirabe K., Fukushima M., Takeshita M., Tanaka H.;
RT   "CEPU-1: an Immunoglobulin Superfamily Molecule, Has Cell Adhesion
RT   Activity and Shows Dynamic Expression Patterns in Chick Embryonic
RT   Spinal Cord.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011810; BAA31514.1; -.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003599; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IG2; 2.
DR   SMART; SM00409; IG2; 2.
DR   SMART; SM00410; IG_like; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   344 AA;  37613 MW;  22CAA8F526A6B57E CRC64;
```

Query Match   68.2%; Score 127.5; DB 13; Length 344;
Best Local Similarity   25.4%; Pred. No. 0.0023;
Matches   68; Conservative   41; Mismatches   98; Indels   61; Gaps   15;

```
Qy    4 VIFLHGSG----SGNEVI-EGPQNATVLKGSQARFNCTVSQ----GW--KLIMWALSD 50
Db   21 LILFLVPAGVPVRSGDATFPKAMDNVTVRQGESATLRCSLQ----GW--KLIMWALSD 80

Qy   51 MVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------ 104
Db   81 IHDVDVYDEGPYTCSVQIDNHPKTSR------------------------NSR 127

Qy  105 LHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPD--ISWELGLLVSHSSY 162
Db  128 VH----LIVQVSPKIEIISSDISINEGNVSLTCIA---TGRPDPITTWR---HISPKAV 177

Qy  163 YFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-----IRCPQDTG 217
Db  178 GFISEDEVLE----ITGITRFQSGEYECSAS-NDVAAPVVQQRVKVTVNYPYISDAKSTG 232

Qy  218 GGINIPGVL----SSLPSLGFSLPTWGK 241
```

Right column

```
Db  233 VPVGQKGILMCEASAVPSADFQ---WYK 257
```

RESULT 14

```
ID   Q61987        PRELIMINARY;      PRT;   871 AA.
AC   Q61987;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Nsk2 protein precursor.
GN   MUSK OR NSK2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Myoblast;
RX   MEDLINE=95349951; PubMed=7624144;
RA   Ganju P., Walls E., Brennan J., Reith A.D.;
RT   "Cloning and developmental expression of Nsk2, a novel receptor
RT   tyrosine kinase implicated in skeletal myogenesis.";
RL   Oncogene 11:281-290(1995).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; X86444; CAA60165.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:103581; Musk.
DR   InterPro; IPR002453; Beta_tubulin.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000024; Fz_domain.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003599; Ig_like.
DR   InterPro; IPR003600; Ig_MHC.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF01392; Fz; 1.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD0000001; Euk_pkinase; 1.
DR   SMART; SM00408; IG2; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50038; Fz; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
KW   ATP-binding; Immunoglobulin domain; Serine/threonine-protein kinase;
KW   Signal; Transferase.
FT   SIGNAL        1        21       POTENTIAL.
SQ   SEQUENCE   871 AA;  97047 MW;  F3C53DC6AFE702AB CRC64;
```

Query Match   68.2%; Score 127.5; DB 11; Length 871;
Best Local Similarity   21.7%; Pred. No. 0.0077;
Matches   70; Conservative   41; Mismatches   125; Indels   87; Gaps   16;

```
Qy    4 VIFLHGSG----SGNEVI-EGPQNATVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY 73
Db   15 EVIGPQNATVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY 73

Qy   74 DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY----LTVQVMGELFIPSVNLVV 128
Db  122 KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD------------NALRENSRIAALE- 171

Qy  129 AENEPCEVTC----LPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQ 183
Db  172 -----SGSLRIHNVQKEDAGYRCVAKNSL--GTAYSKLVKLEVELVGRIIRAPESHNV 223

Qy  184 SNGTLTCVAT----WKSLKARKSATVNLTVIR------------CPQDTGGGINIPG----V 225
```

```
Db    278  --GLYTCIATNKHGEKFPSTAKAAATVSIAEWSKSQKDSQGYCAQYRGEGVLMQGPGEKML  335

Qy    226  LSSLPSLGFSLP------------------TWGKVGLGLAGTMLLTPTCTLITIRCCCRRRCCGCN--  273
Db    336  LVFLPTTSHRDPEDAQELLIHTAWNEL-------KAVSPLCRPAAEALLCYHLFLECSPG  388

Qy    274  -----CCCRCC-----FCCR  283
Db    389  VVPTPMPICREYCLAVKELFCAK  411
```

RESULT 15

```
ID    Q61988
AC    Q61988;
DT    01-NOV-1996 (TrEMBLrel. 01, Created)
DT    01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT    01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE    Nsk2 protein precursor.
GN    MUSK OR NSK2.
OS    Mus musculus (Mouse).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX    NCBI_TaxID=10090;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=MYOBLAST;
RX    MEDLINE=95349951; PubMed=7624144;
RA    Ganju P., Walls E., Brennan J., Reith A.D.;
RT    "Cloning and developmental expression of Nsk2, a novel receptor
RT    tyrosine kinase implicated in skeletal myogenesis.";
RL    Oncogene 11:281-290(1995).
CC    -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR    EMBL; X86445; CAA60166.1; -.
DR    HSSP; P11362; 1FGK.
DR    MGD; MGI:103581; Musk.
DR    InterPro; IPR0024453; Beta_tubulin.
DR    InterPro; IPR000719; Euk_pkinase.
DR    InterPro; IPR000024; Fz_domain.
DR    InterPro; IPR003598; Ig_c2.
DR    InterPro; IPR003600; Ig_like.
DR    InterPro; IPR003006; Ig_MHC.
DR    InterPro; IPR002290; Ser_thr_pkinase.
DR    InterPro; IPR001245; Tyr_pkinase.
DR    Pfam; PF01392; Fz; 1.
DR    Pfam; PF00047; Ig; 3.
DR    Pfam; PF00069; pkinase; 1.
DR    PRINTS; PR00109; TYRKINASE.
DR    ProDom; PD000001; Euk_pkinase; 1.
DR    SMART; SM00408; IGc2; 2.
DR    SMART; SM00410; IG_like; 1.
DR    SMART; SM00219; TyrKc; 1.
DR    SMART; SM00038; Fz; 1.
DR    PROSITE; PS50038; PROTEIN_KINASE_ATP; 1.
DR    PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR    PROSITE; PS50011; PROTEIN_KINASE_ST; 1.
DR    PROSITE; PS00108; PROTEIN_KINASE_TYR; 1.
DR    PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR    PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
KW    ATP-binding; Immunoglobulin domain; Serine/threonine-protein kinase;
KW    Signal; Transferase.
FT    SIGNAL        1       21       POTENTIAL.
SQ    SEQUENCE  881 AA;  98435 MW;  EA0D0282E9B28ED7 CRC64;
```

```
Query Match             6.2%;   Score 127.5;   DB 11;   Length 881;
Best Local Similarity  21.7%;    Pred. No. 0.0078;
Matches  70; Conservative 41; Mismatches 125; Indels 87; Gaps 16;
```

```
Qy   15  EVIEGPQNATVLKGSQARFNCTVSQGWK-LIMWALSDMVLSVRMEPIITNDRFTSQRY  73
         ::    :   ::|  :|:|    |   |||  ||| |:::| |  ||:::|||::|||
Db  122  KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD------NALRENSRIAALE-  171

Qy   74  DQGGNFTSEMIIHNVEPSDSGGNIRCSLQNSRLHGSAY----LTVQVMGELFIPSVNLVV  128
         : |||||:  |::|::  |:|||  ::|| :|:|||:        ||:|  |::|: |
```

Search completed: April 28, 2003, 21:07:34
Job time : 42.7435 secs

```
Qy  226  LSSLPSLGFSLP------------------TWGKVGLGLAGTMLLTPTCTLITIRCCCRRRCCGCN--  273
         |:|                           ||:|                ::|
Db  336  LVFLPTTSHRDPEDAQELLIHTAWNEL-------KAVSPLCRPAAEALLCYHLFLECSPG  388

Qy  274  -----CCCRCC-----FCCR  283
         |:      |:|     ||:|
Db  389  VVPTPMPICREYCLAVKELFCAK  411

Qy  278  --GLYTCIATNKHGEKFPSTAKAAATVSIAEWSKSQKDSQGYCAQYRGEGVLMQGPGEKML  335
         |:|  ||| |                         :  | |::|||:|:|: :| | :|:| V
Db  184  SNGTLTCVAT----WKSLKARKSATVNLTVIR------------CPODIGGGINIPG-----  225

Qy  336  LVFLPTTSHRDPEDAQELLIHTAWNEL-------KAVSPLCRPAAEALLCYHLFLECSPG  388
Qy  224  TFGSFVTLRCTEIGTP-----VPTISWIENGNAVSSGSIQESVKDRVIDSRLQLFITKP-  277
Db  129  AENEPCEVTC----LPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQ  183
Db  172  ------SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKLVKLEVEVLGRILRAPESHNV  223
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:　　　　　　April 28, 2003, 23:25:03 ; Search time 3225.61 Seconds
　　　　　　　　　　　　　　　　　　　　　　　　(without alignments)
　　　　　　　　　　　　　　　　　　　　　　10538.186 Million cell updates/sec

Title:　　　　　　　　US-09-729-264-3
Perfect score:　　　1168
Sequence:　　　　　　1 agtgatcatgtggtggcaggag...........gtaatacaactgtagtatag 1168

Scoring table:　　　IDENTITY_NUC
　　　　　　　　　　　Gapop 10.0 , Gapext 1.0

Searched:　　　　　　2054640 seqs, 1455140287B residues

Total number of hits satisfying chosen parameters:　　　　4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
　　　　　　　　　Maximum Match 100%
　　　　　　　　　Listing first 45 summaries

Database :　　　GenEmbl:*

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_vi:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_mus:*
33: em_htg_other:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1168	100.0	1168	6	AX380398	AX380398 Sequence
2	1122.4	96.1	1175	6	AX380396	AX380396 Sequence
3	1119.6	95.9	2051	9	AK092516	AK092516 Homo sapi
4	1119	95.8	1240	6	AX380400	AX380400 Sequence
5	1102.8	94.4	1139	6	AX380402	AX380402 Sequence
6	378.4	32.4	1195	6	AX380404	AX380404 Sequence
7	326.6	28.0	1427742	9	AF121782	AF121782 Homo sapi
8	326.5	28.0	340000	9	HS21C080	AL163280 Homo sapi
9	182.4	15.6	196900	2	AC020851	AC020851 Mus muscu
10	181.8	15.6	754	6	AX380408	AX380408 Sequence
11	181.8	15.6	895	6	AX380406	AX380406 Sequence
12	181.8	15.6	1556	10	BC004806	BC004806 Mus muscu
13	171.4	14.7	175861	2	AC120346	AC120346 Mus muscu
14	100	8.6	199665	2	AC017582	AC017582 Drosophil
15	57.4	4.9	196900	2	AC020851	AC020851 Mus muscu
16	56.6	4.8	427	9	HSU63332	U63332 Human super
17	55.4	4.7	53908	2	AC120145	AC120145 Mus muscu
18	55	4.7	1061	9	HSCAGCTG	AJ000501 Homo sapi
19	54.8	4.7	175861	2	AC120346	AC120346 Mus muscu
20	53	4.5	178052	2	AC118358	AC118358 Rattus no
21	52.8	4.5	138685	2	AC017582	AC017582 Drosophil
22	52.8	4.5	162249	2	AC093502	AC093502 Drosophil
23	52.8	4.5	185404	3	AC104703	AC104703 Drosophil
24	52.8	4.5	341319	3	AE003536	AE003536 Drosophil
25	51.8	4.4	397	6	BD010479	BD010479 Diagnosti
26	51.8	4.4	85584	2	AC099238	AC099238 Rattus no
27	50.8	4.3	162971	2	AC103439	AC103439 Rattus no
28	50.6	4.3	147185	9	AC020589	AC020589 Homo sapi
29	50.6	4.3	165675	2	AC113925	AC113925 Rattus no
30	50.6	4.3	195460	2	AC129068	AC129068 Rattus no
31	50	4.3	94674	2	AC128466	AC128466 Rattus no
32	50	4.3	179494	2	AC113770	AC113770 Rattus no
33	50	4.3	185487	2	AC108346	AC108346 Rattus no
34	49.8	4.3	299	9	HSA131896	AJ131896 Homo sapi
35	49.8	4.3	161170	2	AC129166	AC129166 Rattus no
36	49.4	4.2	49845	2	AC100974	AC100974 Rattus no
37	49.4	4.2	160650	3	AC008188	AC008188 Drosophil
38	49.4	4.2	160885	2	AC018184	AC018184 Drosophil
39	49.4	4.2	162384	2	AC007840	AC007840 Drosophil
40	49.4	4.2	172334	2	AC122602	AC122602 Rattus no
41	49.4	4.2	173373	3	AC099016	AC099016 Drosophil
42	49.4	4.2	308854	3	AE003797	AE003797 Drosophil
43	49.4	4.2	312634	2	AC095096	AC095096 Rattus no
44	49.4	4.2	313820	2	AC115350	AC115350 Rattus no
45	49.2	4.2	233337	2	AC126948	AC126948 Rattus no

ALIGNMENTS

RESULT 1
AX380398
LOCUS　　　　AX380398　　　　　　　　1168 bp　　DNA　　　linear　　PAT 18-MAR-2002
DEFINITION　Sequence 3 from Patent WO200710.
ACCESSION　 AX380398
VERSION　　　AX380398.1　　GI:19575328
KEYWORDS　　 .
SOURCE　　　　human.
ORGANISM　　 Homo sapiens
　　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　　1
AUTHORS　　　Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
TITLE　　　　B7-like molecules and uses thereof
JOURNAL　　　Patent: WO 0200710-A 3 03-JAN-2002;

FEATURES   Location/Qualifiers
     source   1..1168
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS   8..1168
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="CAD28217.1"
                     /db_xref="GI:19575329"
                     /translation="MVAGAMENRDPGCSGSGNEVIEGPQNARVLKGSQARFNCTVSQG
                     WKLIMWALSDMVVLSVRPMEPIITNDRFTSQRYDQQGNFTSEMIIHNVEPSDSGNIRC
                     SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTMLPDISWELGLL
                     VSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQ
                     DTGGINIPGVLSSLPFSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCCRRKCCGC
                     NCCRCCRCRRKRGFRIQFQFQKKSEKETKNKETETESSGNENSGYNSDSEQKTTDTASLP
                     PKSCESSDPEQRNSSCGPPHQRADQRPPRASHPQASFNLASPEKVSNTTVV"

ORIGIN

BASE COUNT   299 a   302 c   288 g   279 t

Query Match   100.0%;   Score 1168; DB 6;   Length 1168;
Best Local Similarity 100.0%;   Pred. No. 2.4e-270;
Matches 1168; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 AGTGATCATGTGGGTGGCAGGAGCCATGGAAAATAGAGACCCACCGGGTTCTGGTAA 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 AGTGATCATGTGGGTGGCAGGAGCCATGGAAAATAGAGACCCACCGGGTTCTGGTAA 60

Qy   61 TGAAGTCATAGAAGGGCCCCCAAAAATGCAAGAGTCCTGAAAGGGCTCGCTTCAA 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 TGAAGTCATAGAAGGGCCCCCAAAAATGCAAGAGTCCTGAAAGGGCTCGCTTCAA 120

Qy 121 CTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCTCTCAGTGACATGGTGGTGCT 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCTCTCAGTGACATGGTGGTGCT 180

Qy 181 AAGCGGTCAGGCCCATGGAGCCCCATCATCACCAATGACCGGTTCACCTCTCAGAGGTACGA 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 AAGCGGTCAGGCCCATGGAGCCCCATCATCACCAATGACCGGTTCACCTCTCAGAGGTACGA 240

Qy 241 CCAGGGCGGGAACTTCACCTCGGAAGATGATCATCACAATGTGGAGCCCAGTGAATCGGG 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 CCAGGGCGGGAACTTCACCTCGGAAGATGATCATCACAATGTGGAGCCCAGTGAATCGGG 300

Qy 301 GAACATCAGATGCAGCCTCCAGAACAAGTCGCTGCATGGGAGCTACTTACCCGTCCA 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 GAACATCAGATGCAGCCTCCAGAACAAGTCGCTGCATGGGAGCTACTTACCCGTCCA 360

Qy 361 AGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATCTGTTGTAGTCGCTGAGAATGAACCTTG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATCTGTTGTAGTCGCTGAGAATGAACCTTG 420

Qy 421 TGAAGTTACTTGTCTACCCTCACACTGGCACCTGGCTCCCGGATATTCCTGGGAGCTCGG 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 TGAAGTTACTTGTCTACCCTCACACTGGCACCTGGCTCCCGGATATTCCTGGGAGCTCGG 480

Qy 481 TCTCCTGGTCAGCCAATTCAAGCTATTATTTGTTCCGGAGCCCAGGACTTCAAAGTGC 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 TCTCCTGGTCAGCCAATTCAAGCTATTATTTGTTCCGGAGCCCAGGACTTCAAAGTGC 540

Qy 541 AGTGAGCATCCTGGCTCTGACCCCACAGGACAAATGGGACTTTGACTTGCGTGGCTACTG 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 AGTGAGCATCCTGGCTCTGACCCCACAGGACAAATGGGACTTTGACTTGCGTGGCTACTG 600

Qy 601 GAAGAGCCTGAAAGCCCGAAGTCTGCAACTGTAAAATCTCACTGTGATTCGGTGTCCCCA 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 GAAGAGCCTGAAAGCCCGAAGTCTGCAACTGTAAAATCTCACTGTGATTCGGTGTCCCCA 660

Qy 661 AGACACTGGAGGTGGTTAATTAATTCCAGTGTATTACCGAGTTTAGGTTT 720
        ||||||||||||||||||||||||||||||||||||||||||
Db 661 AGACACTGGAGGTGGTTAATTAATTCCAGTGTATTACCGAGTTTAGGTTT 720

Qy 721 TTCATTGCCTACTTGGGCAAAAGTTGGACTTAGCAGGCACCATGCTTCTGACGGCC 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 2
AX380396

LOCUS   AX380396   1175 bp   DNA   linear   PAT 18-MAR-2002
DEFINITION Sequence 1 from Patent WO0200710.
ACCESSION   AX380396
VERSION   AX380396.1   GI:19575326
KEYWORDS
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE   B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 1 03-JAN-2002;
            Amgen, Inc. (US)
FEATURES   Location/Qualifiers
     source   1..1175
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS   27..1175
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="CAD28216.1"
                     /db_xref="GI:19575327"
                     /translation="MGLVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLI
                     MWALSDMVVLSVRPMEPIITNDRFTSQRYDQQGNFTSEMIIHNVEPSDSGNIRCSLQN
                     SRLHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHS
                     SYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGG
                     INIPGVLSSLPFSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCRRCGCNCCRCCF
                     CCRRKRGFRIQFQFQKKSEKETKNKETETESSGNENSGYNSDSEQKTTDTASLPPKSC
                     ESSDPEQRNSSCGPPHQRADQRPPRASHPQASFNLASPEKVSNTTVV"
ORIGIN

BASE COUNT   295 a   307 c   286 g   287 t

Query Match   96.1%;   Score 1122.4; DB 6;   Length 1175;
Best Local Similarity 99.5%;   Pred. No. 2.0e-259;
Matches 1126; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 721 TTCATTGCCTACTTGGGCAAAAGTTGGACTTAGCAGGCACCATGCTTCTGACGGCC 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 721 TTCATTGCCTACTTGGGCAAAAGTTGGACTTAGCAGGCACCATGCTTCTGACGGCC 780

Qy 781 GACGGTGACTCTTACAATACGGCTGCTGCTGCCGCGGCGTCGTGCGTGTTGGCTGCAACTG 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 781 GACGGTGACTCTTACAATACGGCTGCTGCTGCCGCGGCGTCGTGCGTGTTGGCTGCAACTG 840

Qy 841 CTGCTGCCGGTGGTGTTCTGCGTGTAGAAGAAAGAGGAGATTTCGTATTCAATTTCAAAA 900
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 841 CTGCTGCCGGTGGTGTTCTGCGTGTAGAAGAAAGAGGAGATTTCGTATTCAATTTCAAAA 900

Qy 901 GAAATCTGAAAAAGAGAGAAGACAAAAAAGAACTGACGAAAAGTGGAAATGAAAAACTC 960
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 901 GAAATCTGAAAAAGAGAGAAGACAAAAAAGAACTGACGAAAAGTGGAAATGAAAAACTC 960

Qy 961 CGGCTACAATTCAGATGACAAAAGAACCCACAGACACCGCGCTTCTCTCCCCAAAATCCTG 1020
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 961 CGGCTACAATTCAGATGACAAAAGAACCCACAGACACCGCGCTTCTCTCCCCAAAATCCTG 1020

Qy 1021 TGAATCCAGTGATCCTGACAAAAGAGTAGTGTGGCCCTCTCACCAGCGCGGGCTGA 1080
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1021 TGAATCCAGTGATCCTGACAAAAGAGTAGTGTGGCCCTCTCACCAGCGCGGGCTGA 1080

Qy 1081 TCAACGTCCACCCAGGCCAGCAAGTCATCCACACAGCAGTCCTTCTTTTAATCTGGCCAGTCCTGA 1140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1081 TCAACGTCCACCCAGGCCAGCAAGTCATCCACACAGCAGTCCTTCTTTTAATCTGGCCAGTCCTGA 1140

Qy 1141 GAACGTCAGTAAATACAAACTGTAGTATAG 1168
        ||||||||||||||||||||||||||||||
Db 1141 GAAGGTCAGTAAATACAAACTGTAGTATAG 1168

Qy   37 CCCACCCGGTTCTGGGTCTTGGTAATAGAGACGGCCCCCAAAAATGCAAGAGTCCT 96
        || ||||||||||||||||||||||||||||||||||||||||||||||||||

```
RESULT 3
AK092516
LOCUS       AK092516              2051 bp    mRNA    linear   PRI 15-JUL-2002
DEFINITION  Homo sapiens cDNA FLJ35197 fis, clone PLACE6017788, highly similar
            to IGSF5.
ACCESSION   AK092516
VERSION     AK092516.1  GI:21751130
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE6
            clone:PLACE6017788.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
            and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2051)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES             Location/Qualifiers
     source          1..2051
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="PLACE6017788"
                     /tissue_type="placenta"
                     /clone_lib="PLACE6"
                     /note="cloning vector: pME18SFL3"
BASE COUNT      580 a    463 c    474 g    534 t
ORIGIN
```

```
Query Match            95.9%;  Score 1119.6;  DB 9;  Length 2051;
Best Local Similarity  99.6%;  Pred. No. 1.1e-258;
Matches 1122;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy         44 CCTCCACGGTTCTGGGTCTGGTAATGAAGTTCATAGAAGGCCCCAGAAATGCAACAGTCCT 103
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         44 CCTCCACGGTTCTGGGTCTGGTAATGAAGTTCATAGAAGGCCCCAGAAATGCAACAGTCCT 103

Qy         97 GAAAGGGGCTCCCAGGCTCCGCTTCAACTGCACCAGGGCTGGGAAGCTCATCATGTG 156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        104 GAAAGGGGCTCCCAGGCTCCGCTTCAACTGCACCAGGGCTGGGAAGCTCATCATGTG 163

Qy        157 GGCTCTCAGTGACATGGTGGTTGGCTAAGCGTGCTCAGGGCCCATGAGCGCCATCATCACCAATGA 216
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        164 GGCTCTCAGTGACATGGTGGTTGGCTAAGCGTGCTCAGGGCCCATGAGCGCCATCATCACCAATGA 223

Qy        217 CCGCTTCACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGAGATGATCATCCA 276
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        224 CCGCTTCACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGAGATGATCATCCA 283

Qy        277 CAATGTGGAGCCCAGTGATTCGGGGAACATCAGATGCAGCTTCCAGAACAGTCGCCTGCA 336
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        284 CAATGTGGAGCCCAGTGATTCGGGGAACATCAGATGCAGCTTCCAGAACAGTCGCCTGCA 343

Qy        337 TGGATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATCT 396
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        344 TGGATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATCT 403

Qy        397 TGTAGTCGCTGAGAATGAAGCCTTGTGAAGTTGTTTACTTTGTTACCCTCACACTGGACCTGGCT 456
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        404 TGTAGTCGCTGAGAATGAAGCCTTGTGAAGTTGTTTACTTTGTTACCCTCACACTGGACCTGGCT 463

Qy        457 CCGGATAATTCCTGGGGAGCTCGGTCTCCTGGTCATTCAAGCTATTATTTTTTGTTCC 516
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        464 CCGGATAATTCCTGGGGAGCTCGGTCTCCTGGTCATTCAAGCTATTATTTTTTGTTCC 523

Qy        517 GGAGCCCAGCGACCTTCAAAGTGCAGTGAGCAATCCTCGGCTCTGACCCCACACAGCAATGG 576
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        524 GGAGCCCAGCGACCTTCAAAGTGCAGTGAGCAATCCTCGGCTCTGACCCCACACAGCAATGG 583

Qy        577 GACTTTGACTTGCGTGGCTACCTGGAAGAGCCGAAGTCCTGCAACTGTAAA 636
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        584 GACTTTGACTTGCGTGGCTACCTGGAAGAGCCGAAGTCCTGCAACTGTAAA 643

Qy        637 TCTCACTGTGTGATTCGGTGTCCCAAGACACTGGAGGTGGTATTATTAATATTCCAGGTGTATT 696
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        644 TCTCACTGTGATTCGGTGTCCCAAGAACACTGGAGGTGGTATTATTAATATTCCAGGTGTATT 703

Qy        697 ATCAAGTTACCGAGTTTAGGTTTTTTCATTGCCTACTTGGGGCAAAGTTGGACTTGGACT 756
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        704 ATCAAGTTACCGAGTTTAGGTTTTTTCATTGCCTACTTGGGGCAAAGTTGGACTTGGACT 763

Qy        757 AGCAGGCACCATGCTTCTGACGGCCGACGTGTACTCTTACAATACCGCTGCTGCTGCCG 816
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        764 AGCAGGCACCATGCTTCTGACGGCCGACGTGTACTCTTACAATACCGCTGCTGCTGCCG 823

Qy        817 CCGTCGTTGTTGTGGCTGCAACTGCTGCTGGTTGGTTGTTGTTTCTGGCTGTGAGAGAAAAG 876
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        824 CCGTCGTTGTTGTGGCTGCAACTGCTGCTGGTTGGTTGTTGTTTCTGGCTGTGAGAGAAAAG 883

Qy        877 AGGATTTCGTAATTCAATTTGAAAAGAAAATCTGAAAAGAGAGAAGAGAGACAAACTGA 936
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        884 AGGATTTCGTAATTCAATTTGAAAAGAAAATCTGAAAAGAGAGAAGAGAGACAAAACTGA 943

Qy        937 GACAGAAAGTGGAAAATGAAAACTCCGGCTACAATTCAGATGAGAACAAAGACCACAGACAC 996
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        944 GACAGAAAGTGGAAAATGAAAACTCCGGCTACAATTCAGATGAGAACAAAGACCACAGACAC 1003

Qy        997 CGGTTCTCTCGTCTCCCCAAAATCCTGTGATCTGAACAAGACAAGTAGCTG 1056
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1004 CGGTTCTCTCGTCTCCCCAAAATCCTGTGATCTGAACAAGACAAGTAGCTG 1063

Qy       1057 TGGCCCTCCTCCACCAGCGGCCGGGGCTGATCAACGTCCACCCAGGCC 1116
              |||||||||||||||||||||||||||||||||||||||||||||||
Db       1064 TGGCCCTCCTCCACCAGCGGCCGGGGCTGATCAACGTCCACCCAGGC 1123

Qy       1117 TTCTTTTAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTAGINTAG 1168
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1124 TTCTTTTAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGINTAG 1175
```

```
Qy         43 CGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGCCCCAGAAATGCAAGAGG 102
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        202 CGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGCCCCAGAAATGCAAGAGG 261

Qy        103 CTCCCAGGGCTGCCTTCAACTGCACCCTCTCCCAGGGCTGGAAGCTCATCATGTGGGCTCT 162
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        262 CTCCCAGGGCTGCCTTCAACTGCACCCTCTCCCAGGGCTGGAAGCTCATCATGTGGGCTCT 321

Qy        163 CAGTGACATGGTGGTTGCTAAGCGTGCTCAGGGCCCATGAGCGCCATCATCACCAATGACGCTT 222
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        322 CAGTGACATGGTGGTTGCTAAGCGTGCTCAGGGCCCATGAGCGCCATCATCACCAATGACGCTT 381

Qy        223 CACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATCATCACCAATGT 282
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        382 CACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATCATCACCAATGT 441

Qy        283 GGAGCCCAGTGATTCGGGGAACATCAGATGCAGCCTCCAGAACAGTCGCCTCCATGGATC 342
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        442 GGAGCCCAGTGATTCGGGGGACATCAGACATGCAGCCTCCAGAACAGTCGCCTGCCATGGATC 501
```

```
RESULT 4
LOCUS       AX380400                1240 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 5 from Patent WO0200710.
ACCESSION   AX380400
VERSION     AX380400.1  GI:19575330
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE     B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 5 03-JAN-2002;
```

```
                     Amgen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1240
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             80..1240
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="CAD28218.1"
                     /db_xref="GI:19575331"
                     /translation="MERHLLTVPEAVGSGSGNEVIEGPONATVLKGSOARFNCTVSOG
                     WKLIMWALSDMVLSVRPMEPIITNDRFTSQRYDQGGNFTISMIIHNVEPSDSGNIRC
                     SLQNSRLHGSAYLIVQVMGELFIPSVNLVVAENEPCEVTCLFSHWTRLPDISWELGLL
                     VSHSSYYEVPEPSDLQSAVSILALTPQSNGTLTCVAATWKSLKARKSATVNLIVIRCPQ
                     DTGGGINIPGVLSSLPSLGFSLPTWGKYGLGLAGTMLLTPTQTLTIRCCCRRRCCGC
                     NCCCRCFCCRRRGFRIOFQKKSEKEKNKETETESGNENSGYNSDEQKTTETASLP
                     PKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASFNLASPEKVSNTTVV"

BASE COUNT      318 a    319 c    305 g    298 t
ORIGIN
```

```
Query Match      95.8%;  Score 1119;  DB 6;  Length 1240;
Best Local Similarity    99.1%;
Matches 1125; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
Qy     34 AGACCCACCCGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGCCCCCAAAATGCAAGAGT 93
          |||   ||||||||||||||||||||||||||||||||||||||||||||| |||
Db    106 AGAAGCTGTTAGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGCCCCCAGAGATGCAAGAGT 165

Qy     94 CCTGAAGGGGCTCCCAGGGTCGGTTCACGTTCGGTCGGTGGAAGCTCATCAT 153
          ||||  |||||||||||||||||||||||||||||||||||||||||||||
Db    166 CCTGAAGGGGCTCCCAGGGTCGGTTCACGTTCGGTCGGTGGAAGCTCATCAT 225

Qy    154 GTGGGCTCTCAGTGACAATGGTGGTGCTAAGCGTCAGCGCCCATGGGAGCCCATCATCACCA 213
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    226 GTGGGCTCTCAGTGACAATGGTGGTGCTAAGCGTCAGCGCCCATGGGAGCCCATCATCACCA 285

Qy    214 TGACCGCTTCACCTCTCAGAGGTACGACCAGGCGGGGAACTTCACCTCGGAGATGAATCAT 273
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    286 TGACCGCTTCACCTCTCAGAGGTACGACCAGGCGGGGAACTTCACCTCGGAGATGAATCAT 345

Qy    274 CCACAATGTGGAGCCCAGTGATTCGGGGACATCAGATCAGATGAACAGAGTCGCCT 333
          |||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db    346 CCACAATGTGGAGCCCAGTGATTCGGGGACATCAGATCAGATGAACAGAGTCGCCT 405

Qy    334 GCATGGATCCTTACCTTACCGTCCAAGTTATGGGAGGAGCTGTTCATTCCCAGTGTTAA 393
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    406 GCATGGATCCTTACCTTACCGTCCAAGTTATGGGAGGAGCTGTTCATTCCCAGTGTTAA 465

Qy    394 TCTTGTAGTCGGCTGAGAATGAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACCTG 453
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    466 TCTTGTAGTCGGCTGAGAATGAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACCTG 525

Qy    454 GCTCCCGGATATTTCCTGGGGAGCTCGGTCTCCTGGTCACCCATTCAAGCTATTATTTTGT 513
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    526 GCTCCCGGATATTTCCTGGGGAGCTCGGTCTCCTGGTCACCCATTCAAGCTATTATTTTGT 585

Qy    514 TCCGGAGCCCAGCGACCTTCAAAGTGCAGTGACGATCCTGGCTCTGACCCCACAGAGCAA 573
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    586 TCCGGAGCCCAGCGACCTTCAAAGTGCAGTGACGATCCTGGCTCTGACCCCACAGAGCAA 645

Qy    574 TGGGACTTTGACTTGCGTGGCCTGGCTACCTGGAAGAGCCCGAAGGCCCGCAAGTCTGCCAACTGT 633
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    646 TGGGACTTTGACTTGCGTGGCCTGGCTACCTGGAAGAGCCCGAAGGCCCGCAAGTCTGCCAACTGT 705

Qy    634 AAATCTCACTGTGATTACGGTGTCCCAAGACACTGGAGTGGTATTATTATTCCAGTGT 693
          |||   |||||||||||||||||||||||||||||||||||||||||||||
Db    706 AAATCTCACTGTGATTCGGTGTCCCAAGACACTGGAGTGGTATTATTATTCCAGTGT 765

Qy    694 ATTATCAAGTTTACCGAGTTTAGGTTTTTCATTGCCTACTGGGACTGGG 753
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    766 ATTATCAAGTTTACCGAGTTTAGGTTTTTCATTGCCTACTGGGACTGGG 825

Qy    754 ACTGAGCAGGCAGGCACCAATGCGTTCTGACGGCCGACGTGTACTTCTTACAATACGCTGCTGTGCTG 813
```

```
Qy    343 TGCTTACCTTACCGTCCAAGTTTATGGGAGGAGCTGTTCATTCCCAGTGTTAAATCTTGTAGT 402
Db    502 TGCTTACCTTACCGTCCAAGTTTATGGGAGGAGCTGTTCATTCCCAGTGTTAAATCTTGTAGT 561

Qy    403 CGGCTGAGAATGAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACCTCCCGGA 462
Db    562 CGGCTGAGAATGAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACCTCCCGGA 621

Qy    463 TATTTCCTGGGGAGCTCGGTCTCCTGGTCACCCATTCAAGCTATTATTTTGTTCCGGAGCC 522
Db    622 TATTTCCTGGGGAGCTCGGTCTCCTGGTCACCCATTCAAGCTATTATTTTGTTCCGGAGCC 681

Qy    523 CAGCGACCTTCAAAGTGCAGTGACGATCCTGGCTCTGACCCCACAGAGCAATGGGACTTT 582
Db    682 CAGCGACCTTCAAAGTGCAGTGACGATCCTGGCTCTGACCCCACAGAGCAATGGGACTTT 741

Qy    583 GACTTGCGTGGCCTGGCTACCTGGAAGAGCCCGCAAGTCTGCCAACTGTAAATCTCAC 642
Db    742 GACTTGCGTGGCCTGGCTACCTGGAAGAGCCCGCAAGTCTGCCAACTGTAAAATCTCAC 801

Qy    643 TGTGATTCGGTGTCCCAAGACACTGGAGTGGTATTAATTATTCCAGTGTATTATCAAG 702
Db    802 TGTGATTCGGTGTCCCAAGACACTGGAGTGGTATTAATTATTCCAGTGTATTATCAAG 861

Qy    703 TTTACCGAGTTTAGGTTTTTCATTGCCTACTGGGACTGGGCCAAAGTTGGACTCAGG 762
Db    862 TTTACCGAGTTTAGGTTTTTCATTGCCTACTGGGACTGGGCCAAAGTTGGACTCAGG 921

Qy    763 CACCATGCTTCTCTGACGGCCGACGTGTACTTACTCTTACAATACGCTGCTGTCGCGCGCTCG 822
Db    922 CACCATGCTTCTCTGACGGCCGACGTGTACTTACTCTTACAATACGCTGCTGTCGCGCGCTCG 981

Qy    823 TTGTTGTGGCTGCAACTGCTGTCTGCTCCGGTTGTTTCTGTGTTAGAAGCAAAAAGAGGATT 882
Db    982 TTGTTGTGGCTGCAACTGCTGTCTGCTCCGGTTGTTTCTGTGTTAGAAGCAAAAAGAGGATT 1041

Qy    883 TCGTATTCAAATTTCAAAAGAAAACTGAAAAAAAGAGACACAAACAAAGAAAACTGAGACAGA 942
Db   1042 TCGTATTCAAATTTCAAAAGAAAACTGAAAAAAAGAGACACAAACAAAGAAAACTGAGACAGA 1101

Qy    943 AAGTGGAAATGAAAACTCCGGCTACAATTCAGATGAACAAGTAGTCGTGTGTGGCCC 1002
Db   1102 AAGTGGAAATGAAAACTCCGGCTACAATTCAGATGAACAAGTAGTCGTGTGTGGCCC 1161

Qy   1003 TCTCCCCTCCCAAATCCTGTGAATCCAGTGATCCTGAACAAAGAAACAGTAGCTGTGGCCC 1062
Db   1162 TCTCCCCTCCCAAATCCTGTGAATCCAGTGATCCTGAACAAAGAAACAGTAGCTGTGGCCC 1221

Qy   1063 TCCTCACCAGCGGCGGGCGGGCTGATCAACGTCCACCCAGGCCAGCAAGTCATCCACAGGCTTCTTT 1122
Db   1222 TCCTCACCAGCGGGCGGGCGGGCTGATCAACGTCCACCCAGGCCAGCAAGTCATCCACAGGCTTCTTT 1281

Qy   1123 TAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTAGTAG 1168
Db   1282 TAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTAGTATAG 1327
```

```
RESULT 5
LOCUS       AX380402       1139 bp    DNA      linear   PAT 18-MAR-2002
DEFINITION  Sequence 7 from Patent WO0200710.
ACCESSION   AX380402
VERSION     AX380402.1  GI:19575332
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE     B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 7 03-JAN-2002;
            Amgen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1139
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             1..1134
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD28219.1"
                     /db_xref="GI:19575333"
                     /translation="MVAGAMENRDPGCSSGSGNEVIEGPQNARVLKGSQARFNCTVSQG
                     WKLIMWALSDMVLSVRPMEPIITNDRFTSQRYDQGGNLTSEMIIHNVEPSDSGNIRC
                     SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLL
                     VSHSSYFVFPEPSDLQSAVSIIALTPQSNGTLTVCVATWKSLKARKSATVNLIVIRCPQ
                     DTGGGINIPGVLLSSLPSLGFSPLPTWGKVGLGLAGTMLLTPTCLTLIRCCCRRRCGGC
                     NCCRCFPCCRKRGNLKKRQTKKLRQKVEMKTPATIQMNKRPQTPLLSLPNPVNPV
                     ILNKETVAVALLTSGLINVHPGQQVIHRLLLITWPVLRRSVIQL"
BASE COUNT      290 a    300 c    283 g    266 t
ORIGIN
```

```
Qy       1  ATGGTGGCAGGAGCCATGGAGAATAGAGACCCA...CACCCACCGGTTCTCGGGTTCTGGTAATGAAGTC   67
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1  ATGGTGGCAGGAGCCATGGAGAATAGAGACCCA...CACCCACCGGTTCTCGGGTTCTGGTAATGAAGTC   67

Qy      68  ATAGAGAGGCCCCAAAATGCAAGAGTCCTGAAGGGCTCCCAGGCTCACC                      127
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68  ATAGAGAGGCCCCAAAATGCAAGAGTCCTGAAGGGCTCCCAGGCTCACC                      127
```

```
Qy       8  ATGGTGGCAGGAGCCATGGAGAAATAGAGACCCA     60
            ||||||||||||||||||||||||||||||||||
Db       1  ATGGTGGCAGGAGCCATGGAGAAATAGAGACCCA     60
```

Best Local Similarity 97.9%;
Query Match 94.48; Score 1102.8; DB 6; Length 1139;
Matches 1136; Conservative 0; Mismatches 2; Indels 22; Gaps 1;
Pred. No. 1.2e-254;

```
RESULT 6
LOCUS           AX380404              1195 bp    DNA    linear    PAT 18-MAR-2002
DEFINITION      Sequence 9 from Patent WO0200710.
ACCESSION       AX380404
VERSION         AX380404.1  GI:19575334
KEYWORDS        .
SOURCE          house mouse.
  ORGANISM      Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE       1
  AUTHORS       Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE         B7-like molecules and uses thereof
  JOURNAL       Patent: WO 0200710-A 9 03-JAN-2002;
                Amgen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1195
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             53..1165
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD28220.1"
                     /db_xref="GI:19575335"
                     /translation="MEGSWRDVLAVLVILAQLTASGSSYQIIEGPQNVTVLKDSEAHF
                     NCTVTHGWLLMWTLNQMVVLSLTTGGPIITNNRTYASTNSDSFISELIIHDVQPS
                     DSGSVQCLSQNSHGFGSAFLSVQVMGTLNIPSNLIVTEGEPGYNVKRSLTVVGWTSLPDI
                     SWELEVPVSHSSYNSFLEPGNFMRVLSVLDLTPLGNGTLCVAELKDLQASKSITVNL
                     TVQPPPSIGEGPALPTWAIILLLAVAFSLLLILIIVIIIIFPCCCCASRREKEESTY
                     QNEIRKSANMRTNKAPDETKLKSGKENYGYSSDEAKAQTASLPKSAEVSLPEKRSS
                     SLPYQELNKHQPGPATHPRVSFDIASPQKVRNVTLV"
BASE COUNT        313 a    304 c    291 g    287 t
ORIGIN

Query Match      32.4%;  Score 378.4;  DB 6;  Length 1195;
Best Local Similarity  62.3%;  Pred. No. 2.5e-80;
Matches 701;  Conservative  0;  Mismatches 356;  Indels  69;  Gaps  4;

Qy     42 CCGGTTCTTGGGGTCTGTAAATGAAGTCATGATAGAGAGCCCCCAAAAATGCAAGAGTCCTGAAGG 101
          |||| ||| |||||| ||||| || ||| |||| ||  |||||| |||  ||| ||| |||| |||
Db    108 CAGCTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAATGTACAGTCCTAAAGG 167

Qy    102 GCTCCCAAGGCTCGCTTCAACTGACCGTCTCCCAGGGTGGAGAAGCTCATCATGTGGGCTC 161
          |||  || |||| || |||||| |||| ||||| |||||||| || |||| ||| ||||||
Db    168 ACTCAGAGGCTCACTTCAACTGACCGTGACCGTGAAGCTTCTCATGTGGACTC 227

Qy    162 TCAGTGACACAGGTGGTGGTGCTAAGCGTGCAGCCCATGGAGCCCATCATCACCAATGACCGCT 221
          ||| || || ||||| || |||||| |||||  |||| ||||| ||| |||| ||| |||||
Db    228 TTAACCAAATGGTGGTGCTGAGTCTCACCAACACCACCACCGCT 287

Qy    222 TCACCTCTCAGAGGTTACGACCAGGGCGGGAACTTCACCTCTCGGAGATGAATGATCATCCACAATG 281
          |||||| | ||||| ||| || ||||| |||||||||| | |||| ||||| || ||| |||||
Db    288 TCACCTATCCAGTTAACAGCACACTGACAGTTCAATCTCGGAGTTGATCATCCATCAGTGGGCT 347

Qy    282 TGGAGCCCAGTGATTCGGGGAACATCAGATGCAGCCTCCAGACACAGTCGCGCTGCATGGGAT 341
          || || |||| || ||||| ||||| | || |||| ||| |||| || ||||| ||||||||
Db    348 TGCAGCCCAGTGACTGGTGCTGCTGAATGCAGCCACCGGTGAGAACAGCACCATGGGGTTTGGAT 407

Qy    342 CTGCTTACCTTACCGTCCAGTTATGGGAGAGAGCTGTTCATTCCCAGTGTTAATCTTGTAG 401
          |||| | ||| ||||| |||||| ||| ||||| || | ||  ||| |||| ||||||||
Db    408 CTGCCTTCCTCTCAGTGCAAGTCACCGGACCTGAACATTCCTAGCAACAACCTTATAG 467

Qy    402 TCGCTGACAATGAACCTTGTGTCGAAGTTACTTGTCTACCCTCACACTGACCTGCCTCCCG 461
          || || || |||| |||  ||| || ||| || ||  || ||| || || |||| |||
Db    468 TCACTGAGGGTGAACCCTGTAATGTACCGTGGGCCTGAACCTCCACTCCCG 527

Qy    462 ATATTTCCTGGGAGCTGGAGGTCGGCGTCTCCTGGTCCAGCCATTCAAGCTATTATTTGTTCCGGAGC 521
          |||||||  ||||||||||||  |||||| ||| |||  | ||| ||  ||||| | |||||||||||
Db    528 ATATTCCTGGGAGCTTGGAGGTTCCGTAAGCCATTCGAGTTACAATTCCTTTCTGGGAGC 587
```

```
RESULT 7
LOCUS           AF121782            142742 bp    DNA    linear    PRI 02-FEB-1999
DEFINITION      Homo sapiens chromosome 21q22.3 PAC 206A10, complete sequence.
ACCESSION       AF121782
VERSION         AF121782.1  GI:4210991
KEYWORDS        HTG.
SOURCE          Homo sapiens.
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE       1  (bases 1 to 142742)
  AUTHORS       Taudien,S., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
                Schattevoy,R., Weber,J., Schilling,M., Menzel,U., Yaspo,M.L. and
                Rosenthal,A.
  TITLE         Direct Submission
  JOURNAL       Submitted (21-JAN-1999) Genome Analysis, Institute for Molecular
                Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1..142742
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="PAC 206A10"
     repeat_region   3..118
                     /evidence=not_experimental
                     /rpt_family="MIR"

Qy    522 CCAGCGACCTTCAAAGTGCAGTGAGCAATCCTGGCTCTGTGACCCGAGAGCAATGGGACTT 581
          || || ||| |||| ||||| ||  |||  ||| ||| ||| |||  |||| ||| |||||
Db    588 CGGGCAACTTTATGAGGGTCTTGAGTGTCCTGGGACCTGCACACACTGGGCACCAGGGACCT 647

Qy    582 TGACTTGGGTGGCTACCTGGAGAGCCTGAAGAGCCCGAAGGTCTGCAACTGTAAATCTCA 641
          || ||| |||  || || |||| |||| ||| ||  |||| ||| |||||  |||| |||
Db    648 TGACTTGTGTGGCAGAGCTGAAGGACTTGCAGGCCAGCAAGGTCCTTAACTGTCAACCTGA 707

Qy    642 CTGTGATTCGGTGGTCCCCAAGACACTGGAGGTGCTGTATTAATATTCAGGTGTGTATTATCAA 701
          ||| ||| ||||||| |||||| ||||| ||| || ||||| |||| || || ||||| | ||
Db    708 CTGTGGTTCAGCCTCCACCTGACCTGACCAGTATTGGAACAGGA------------ 744

Qy    702 GTTTACCGAGTTAGGTTTTTTCATTGCCTACTTGGGGGCCAAAGTTGGACTTGGACTAGCAG 761
               ||||  ||| ||||||  |||| |||||| |||| || ||||| | |||| |||| |||
Db    745 ------------AGGCCCAGCACTGCGGACCTGGGCCCCATCATCCGTGGGCAGTGGCGCT 791

Qy    762 GCACCAATGCTTCTGACGCGCCGACCTGTACTCTTACAATACGGCTGTTGCCTTGCCGCCCGTC 821
          |||||| || ||| || || || ||| || | |||| || || ||||| ||||  ||||||
Db    792 TTTCCTTGCTCTTCACTTGTTGACCAGCCTTTAAAATTCTGTTGCTG------------ 843

Qy    822 GTTGTTGTGGGCTGCAACTGCTCTGCTGCGTGTTGTTCTGGTGAGAAGAGGAGGAT 881
          |||||  || ||||||  || || |||| |||||| || || || ||| |||||
Db    844 --------TTGTGCCTTCCAGGAACTCAATAACATCAGCCCGGTCAATAACAACTGTTAATACAACTAGTGATA 898

Qy    882 TTCGTATTCCAATTTCAAAGAAAATCTGAAAGAGAATCTGAAAAGAGAAGAGAGACAAACAAAAGAAGAAAACTAATAG 941
          || || | |||||| ||||| ||||||||| || ||||||  ||||| || || || |||||| |||||||
Db    899 TCTGCAAACATGAGGACAA------------ 944

Qy    942 AAAGTGAAAATGAAAAACTCCGGGCTACAATTCAGATGAACAAAGAGACAACAGAGACAACCGCTT 1001
          ||||| |||||| ||||| | ||| || | ||||| ||||| ||||| |  |||| ||||| | |
Db    945 AAAGTGGAAAAGGAAAAACTACGGGTACAGTTCGGATGAGGCAAGGCTGCACAGAGACTGCAT 1004

Qy   1002 CTCTCCCTCCCAAAATCCTGTGAATCCAGTAGATCCTGAACAAAGAACAAGTAGCTGTTGGCC 1061
          |||||||| ||||||||| |||| ||| |||| ||||| ||||||  ||||| || || |||
Db   1005 CTCTCCCTCCTTAAAATCTGCTGAAGTCACGCCTTCCAGAAAAACGCAGCGAGTA------GCC 1058

Qy   1062 CTCCTCACCAGCGGCGGGCTCGATCACGATCGACCCACCCAGGCCAGCAAGTCAATCCACCAGGGCTTCTT 1121
          |||| || ||  || |||| || | |||| || ||| || || |||| |||| || |||| |||| |||
Db   1059 TTCCTTATCAGGAACTCAATAACATCAGCCCGGTCAATAACAACTAGTGATA------GCC 1118

Qy   1122 TTAATCTGGCCAGTCCAGAAGTTCAGTAATACAACTGTAGTGATA 1167
          ||||| |||||| || ||||| |||||||| ||| |||||| |||
Db   1119 TTGACATCGCCAGTCCTCAGAAGGTTCAGAAATGGACTTTAGTGTA 1164
```

```
repeat_region   complement(140. .415)
                /evidence=not_experimental
                /rpt_family="MLT2E"
repeat_region   complement(668. .881)
                /evidence=not_experimental
                /rpt_family="MER58A"
exon            complement(1078. .1162)
                /note="GRAIL, score = 86.000%, comment = excellent shadow"
                /evidence=not_experimental
repeat_region   1114. .1403
                /evidence=not_experimental
                /rpt_family="MLT1J"
repeat_region   1481. .1603
                /evidence=not_experimental
                /rpt_family="L2"
exon            1576. .1710
                /evidence=not_experimental
                /note="GRAIL, score = 66.18"
exon            1576. .1710
                /note="MZEF, score = 99.000%, comment = excellent"
                /evidence=not_experimental
repeat_region   complement(1693. .1741)
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   complement(1780. .2326)
                /evidence=not_experimental
                /rpt_family="MLT1E"
exon            complement(1834. .1955)
                /note="GRAIL, score = 50.000%, comment = good shadow"
                /evidence=not_experimental
repeat_region   2629. .2831
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   3012. .3122
                /evidence=not_experimental
                /rpt_family="MLT1J"
repeat_region   complement(3502. .3540)
                /evidence=not_experimental
                /rpt_family="MLT1J"
repeat_region   3123. .3478
                /evidence=not_experimental
                /rpt_family="THE1B"
exon            3479. .3618
                /evidence=not_experimental
                /note="GRAIL, score = 65.000%, comment = good shadow"
exon            4674. .4781
                /evidence=not_experimental
                /note="GRAIL, score = 42.000%, comment = marginal"
exon            4712. .4749
                /evidence=not_experimental
                /note="Xpound exon prediction, score = 70% (0%)"
repeat_region   5385. .6104
                /evidence=not_experimental
                /rpt_family="HERVL"
exon            6682. .6751
                /note="MZEF, score = 95.1%"
                /evidence=not_experimental
exon            6746. .6751
                /note="GenScan, score = 1.838, comment = Initial_exon 6 bp
                frame: 1 phase: 0"
repeat_region   6901. .7339
                /evidence=not_experimental
                /rpt_family="MLT1C"
repeat_region   7347. .7457
                /evidence=not_experimental
                /rpt_family="LTR16C"
repeat_region   complement(7561. .7696)
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   complement(7819. .8220)
                /evidence=not_experimental
                /rpt_family="MLT1J"

exon            8097. .8119
                /note="GRAIL, score = 45.000%, comment = marginal"
                /evidence=not_experimental
repeat_region   complement(8738. .9232)
                /evidence=not_experimental
                /rpt_family="MLT1C"
exon            complement(9273. .9420)
                /note="GRAIL, score = 71.000%, comment = good"
                /evidence=not_experimental
exon            9290. .9371
                /evidence=not_experimental
                /note="GRAIL, score = 81.000%, comment = excellent shadow"
repeat_region   9966. .10041
                /evidence=not_experimental
                /note="homology = 100.00%, score = 38, counts = 2"
                /rpt_unit=ttcttgcacacttcctcgttggagggtcggtttatcgc
                /rpt_type=tandem
exon            10244. .10664
                /evidence=not_experimental
                /note="MZEF, score = ..."
repeat_region   10665. .11028
                /evidence=not_experimental
                /rpt_family="MLT1E"
repeat_region   11318. .11417
                /evidence=not_experimental
                /rpt_family="L1MC/D"
exon            complement(11525. .11777)
                /note="MZEF, score = 79.4%"
                /evidence=not_experimental
repeat_region   11503. .11818
                /evidence=not_experimental
                /rpt_family="MLT1C"
repeat_region   12902. .13180
                /evidence=not_experimental
                /rpt_family="AluSg"
exon            complement(13057. .13161)
                /note="MZEF, score = 76.1%"
                /evidence=not_experimental
repeat_region   complement(13226. .13364)
                /evidence=not_experimental
                /rpt_family="L1MC2"
repeat_region   complement(13613. .13908)
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   complement(14317. .14614)
                /evidence=not_experimental
                /rpt_family="LTR37A"
exon            complement(14962. .15036)
                /note="MZEF, score = 93%"
                /evidence=not_experimental
repeat_region   15255. .15299
                /evidence=not_experimental
                /rpt_family="L1PA6"
exon            complement(15469. .15497)
                /note="Xpound exon prediction, score = 62% (0%)"
                /evidence=not_experimental
repeat_region   16045. .16768
                /evidence=not_experimental
                /rpt_family="MER21A"
exon            16555. .16655
                /note="GRAIL, score = 54.000%, comment = good"
                /evidence=not_experimental
repeat_region   16823. .16850
                /evidence=not_experimental
                /rpt_family="MER34"
repeat_region   complement(16851. .17208)
                /evidence=not_experimental
                /rpt_family="THE1A"
repeat_region   17209. .17632
                /evidence=not_experimental
```

```
LOCUS       HS21C080              340000 bp    DNA     linear   PRI 24-MAY-2000
DEFINITION  Homo sapiens chromosome 21 segment HS21C080.
ACCESSION   AL163280 AP001735 BA000005
VERSION     AL163280.2  GI:7717369
KEYWORDS    human.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
```

            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
```
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
            Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
            Group, * Institute of Molecular Biotechnology, Genome Research
            Keio University School of Medicine, Dept. of Molecular Biology, *
            GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
            Genetics (addresses see below)
  COMMENT   The Chromosome 21 Mapping and Sequencing Consortium consists of
            * RIKEN Genomic Sciences Center, Human Genome Research Group,
            Sagamihara 228-8555, Japan,
            * e.mail: sakaki@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Max-Planck Institute for Molecular Genetics,
            Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/.
            and
            * Kelo University School of Medicine, Dept. of Molecular Biology,
            Tokyo 160-8582, Japan,
            * e.mail: shimizu@dmb-med.keio.ac.jp
            * URL: http://adenine.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: gsc1-submit@genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
FEATURES             Location/Qualifiers
     source          1..340000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
     source          <1..125946
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="P70I24"
                     /clone_lib="RPCI1,3-5 PAC library"
                     /note="Accession No. AF064860"
     source          224137..>340000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="P206A10, 3' partial"
                     /clone_lib="RPCI1,3-5 PAC library"
                     /note="Accession No. AF121782"
     source          286628..>340000
```

```
     repeat_region   complement(17633..17969)
                     /rpt_family="MER34"
     repeat_region   complement(18004..18334)
                     /rpt_family="AluSg1"
     repeat_region   complement(18335..18632)
                     /rpt_family="LTR17"
     repeat_region   complement(1852..18420
                     /rpt_family="HERV17"
     exon            1852..18420
                     /evidence=not_experimental
                     /note="GRAIL, score = 51.00%, comment = good shadow"
     exon            complement(1837..18584)
                     /evidence=not_experimental
                     /note="GRAIL, score = 66.00%, comment = good"
     repeat_region   complement(18643..24647)
                     /rpt_family="HERV17"
     exon            complement(19228..19361)
                     /evidence=not_experimental
                     /note="MZEF, score = 92.8%"
     exon            20900..21135
                     /evidence=not_experimental
                     /note="MZEF, score = 51.4%"
     exon            21150..21275
                     /evidence=not_experimental
                     /note="xpound exon prediction, score = 74% (0%)"
```

```
RESULT 8
Query Match   28.0%;  Score 326.6; DB 9; Length 142742;
Best Local Similarity   97.4%;  Pred. No. 9.4e-68;
Matches 332; Conservative   0; Mismatches 9; Indels   0; Gaps   0;

Qy       27 AAAATAAGAGACCCACCCGGTGTCTGGGTCTCGGTAATGAAGTCATAGAAGGCCCCCAAAATG 86
            |||||||||||| |||||||||||||||||||||||||||||||||| ||||||||||||||
Db    44257 AAAATTGGCTACTTCCAGGTTCTGGGTCTCGGTAATGAAGTCATAGAAGGCCCCCAAAATG 44316

Qy       87 CAAGAGTCCTGAAGGCGCTCCCAGGCGCTGCGTTCAACTGCACCGTCTCCCAGGCGCTGGAAGC 146
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    44317 CAAGAGTCCTGAAGGCGCTCCCAGGCGCTGCGTTCAACTGCACCGTCTCCCAGGCGCTGGAAGC 44376

Qy      147 TCATCAATGTGGGCTCTCGATGGACAATGGTGTGGTGTCAGGCCCATGGAGCCCATCA 206
            |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    44377 TCATCAATGTGGGCTCTCAGTGACAATGGTGGTGTTCAGGCCCATGGAGCCCATCA 44436

Qy      207 TCACCAATGACCGCTTCACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGA 266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    44437 TCACCAATGACCGCTTCACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGA 44496

Qy      267 TGATCATCCACAAATGTGGAGCCCCAGTGATTCGGGAGGACCACCTCCAGAACA 326
            ||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    44497 TGATCATCCACAAATGTGGAGCCCCAGTGATTCGGGAGAACATCAGATGCAGCCTCCAGAACA 44556

Qy      327 GTCGCCTGCATGGATCTGCTTACCTTACCGTCCAAGTTATG 367
            ||||||||||||||||||||||||||||||||||||  | |
Db    44557 GTCGCCTGCATGGATCTGCTTACCTTACCGTCCAAGGTGTG 44597
```

```
          /organism="Homo sapiens"
          /db_xref="taxon:9606"
          /chromosome="21"
          /map="21q22.3"
          /clone="BAC-291B3, 3' partial"
          /clone_lib="BAC library"
          /note="Accession No. AF064857"
repeat_region   1088..1566
          /note="L1M4"
          /rpt_family="LINE/L1"
          /rpt_type=DISPERSED
repeat_region   2398..2495
          /note="(TCTA)n"
          /rpt_family="Simple_repeat"
          /rpt_type=TANDEM
repeat_region   complement(2496..2878)
          /note="THE1C"
          /rpt_family="LTR/MaLR"
          /rpt_type=DISPERSED
repeat_region   2879..2931
          /note="(TCTA)n"
          /rpt_family="Simple_repeat"
          /rpt_type=TANDEM
repeat_region   2932..3004
          /note="(TCCA)n"
          /rpt_family="Simple_repeat"
          /rpt_type=TANDEM
repeat_region   3005..3074
          /note="(TCCA)n"
          /rpt_family="Simple_repeat"
          /rpt_type=TANDEM
gene            3473..19157
          /gene="SH3BGR"
mRNA            join(<3473..3565,12649..12678,15331..15426)
          /gene="SH3BGR"
          /partial
CDS             join(3473..3565,12649..12678,15331..15426)
          /gene="SH3BGR"
          /note="Accession No. X93498"
          /codon_start=1
          /product="21-Glutamic Acid Rich protein 21-GARP"
          /protein_id="CAB90445.1"
          /db_xref="GI:7717370"
          /translation="GSEKAEGGETEAQKEGSEDVGNLPEAQEKNEEGEETATETETEE
          IAMEGAEGEAEEEETAEGEPGEDEDDS"
exon            3473..3565
          /gene="SH3BGR"
          /number=4
repeat_region   5542..5829
          /note="AluSx"
          /rpt_family="SINE/Alu"
          /rpt_type=DISPERSED
repeat_region   complement(6887..7067)
          /note="L1MD3"
          /rpt_family="LINE/L1"
          /rpt_type=DISPERSED
repeat_region   complement(8482..8762)
          /note="AluJo"
          /rpt_family="SINE/Alu"
          /rpt_type=DISPERSED
repeat_region   complement(8865..8964)
          /note="AluJo"
          /rpt_family="SINE/Alu"
          /rpt_type=DISPERSED
repeat_region   9131..9424
          /note="AluSq"
          /rpt_family="SINE/Alu"
          /rpt_type=DISPERSED
repeat_region   complement(9723..9882)
          /note="L2"
          /rpt_family="LINE/L2"
          /rpt_type=DISPERSED
repeat_region   complement(9989..10077)
          /note="L2"
          /rpt_family="LINE/L2"
          /rpt_type=DISPERSED
repeat_region   complement(10078..10429)
          /note="THE1B"
          /rpt_family="LTR/MaLR"
          /rpt_type=DISPERSED
repeat_region   complement(10430..10651)
          /note="L2"
          /rpt_family="LINE/L2"
          /rpt_type=DISPERSED
repeat_region   10980..11000
          /note="(TTTA)n"
          /rpt_family="Simple_repeat"
          /rpt_type=TANDEM
repeat_region   complement(11002..11282)
          /note="AluSp"
          /rpt_family="SINE/Alu"
          /rpt_type=DISPERSED
repeat_region   complement(11315..11556)
          /note="AluJo"
          /rpt_family="SINE/Alu"
          /rpt_type=DISPERSED
repeat_region   11702..11753
          /note="MIR"
          /rpt_family="SINE/MIR"
          /rpt_type=DISPERSED
exon            12649..12678
          /gene="SH3BGR"
          /number=5
repeat_region   complement(13115..13425)
          /note="AluSx"
          /rpt_family="SINE/Alu"
          /rpt_type=DISPERSED
repeat_region   complement(13598..13892)
          /note="AluY"
          /rpt_family="SINE/Alu"
          /rpt_type=DISPERSED
repeat_region   14013..14131
          /note="(CA)n"
          /rpt_family="Simple_repeat"
```

```
Query Match      28.0%;  Score 326.6;  DB 9;  Length 340000;
Best Local Similarity   97.4%;  Pred. No. 9.8e-68;
Matches  332;  Conservative  0;  Mismatches  9;  Indels  0;  Gaps  0;

Qy       27 AAAATAGAGACCCACCCCGTTCTGGGGTCTGGGTAATGAAGTCATAGAAGGCCCCCAAAAATG   86
            ||| ||| ||||||||||| |||||||||||||||||||||||||||||||||||||| ||||
Db   268437 AAAATTGGCTACCTTCCAGGTTCTTGGGGTCTGGGTAATGAAGTCATAGAAGGCCCCCAAAATG  146

Qy       87 CAAGAGTCCTGAAGGGCTCCCAGGCCTCGCTTCAACTGCCACCGTCTCCCAGGCCTGGAAGC  146
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   268497 CAAGAGTCCTGAAGGGCTCCCAGGCCTCGCTTCAACTGCCACCGTCTCCCAGGCCTGGAAGC  268556

Qy      147 TCATCATGTGGGCGTCTCAGTGACATGGTGGTGTGCTTAAGCCGTCAGGCCCATGGAGCCCATCA  206
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   268557 TCATCATGTGGGCGTCTCAGTGACATGGTGGTGTGCTTAAGCCGTCAGGCCCATGGAGCCCATCA  268616

Qy      207 TCACCAAATGACCGCGTTCACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGAGA  266
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   268617 TCACCAAATGACCGCGTTCACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGAGA  268676

Qy      267 TGATCATCCACAAATGTTGGAGCCCAGTGATTCGGGGAGAACATCAGATGCACGCCTCAGAACA  326
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   268677 TGATCATCCACAAATGTTGGAGCCCAGTGATTCGGGGAGAACATCAGATGCACGCCTCAGAACA  268736

Qy      327 GTCGCCTTGCATCGATCTGCTTACCTTACCGTCCAAGTTATG  367
            ||||||| |||| ||||||||||||||||||||||||| | ||
Db   268737 GTCGCCTCGCATGGATCTGCTTACCTTACCGTCCAAGGTGTG  268777

RESULT 9
AC020851
```

```
LOCUS       AC020851              196900 bp    DNA     linear   HTG 15-JUL-2000
DEFINITION  Mus musculus clone RP21-467L12, WORKING DRAFT SEQUENCE, 33
            unordered pieces.
ACCESSION   AC020851
VERSION     AC020851.2  GI:9211212
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 196900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 15, 2000 this sequence version replaced gi:6666457.
            -------------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            Project Information
            Center Project Name: 1437240
            Center clone name: RPCI-21_467L12
            ------------
            Summary Statistics
            Consensus quality: 164618 bases at least Q40
            Consensus quality: 180887 bases at least Q30
            Consensus quality: 183777 bases at least Q20
            Estimated insert size: 147000; pulse field gel estimation
            Estimated insert size: 19370; sum-of-contigs estimation
            Quality coverage: 5.67 in Q20 bases; pulse field gel estimation
            Quality coverage: 4.3 in Q20 bases; sum-of-contigs estimation
            NOTE: This is a 'working draft' sequence. It currently
            consists of 33 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
                 1     1541:  contig of 1541 bp in length
              1542     1641:  gap of unknown length
              1642     3492:  contig of 1851 bp in length
              3493     3592:  gap of unknown length
              3593     4738:  contig of 1146 bp in length
              4739     4838:  gap of unknown length
              4839     7248:  contig of 2410 bp in length
              7249     7348:  gap of unknown length
              7349     9178:  contig of 1830 bp in length
              9179     9278:  gap of unknown length
              9279    11493:  contig of 2215 bp in length
             11494    11593:  gap of unknown length
             11594    13607:  contig of 2014 bp in length
             13608    13707:  gap of unknown length
             13708    15946:  contig of 2239 bp in length
             15947    16046:  gap of unknown length
             16047    18392:  contig of 2346 bp in length
             18393    18492:  gap of unknown length
             18493    21231:  contig of 2739 bp in length
             21232    21331:  gap of unknown length
             21332    24662:  contig of 3331 bp in length
             24663    24762:  gap of unknown length
             24763    27140:  contig of 2378 bp in length
             27141    27240:  gap of unknown length
             27241    29950:  contig of 2710 bp in length
             29951    30050:  gap of unknown length
             30051    32245:  contig of 2195 bp in length
             32246    32245:  gap of unknown length
```

```
FEATURES             Location/Qualifiers
     source          1..196900
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP21-467L12"
                     /clone_lib="RPCI mouse PAC library 21"
BASE COUNT    51456 a  43823 c  44608 g  53666 t  3347 others
ORIGIN
```

```
Qy         37 CCCACCCGGTTCTTGGGGTTCTGGTGAATGAAGTCATAGAAGGCCCCAAAATGCAAGAGTCCT 96
              ||  |||||||||||| ||  ||||||||||||||||||||||||||||||||||||||| |||
Db      67439 CCTTCTAACTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAAATGTAACAGTCCTT 67498

Qy         97 GAAAGGGGTCCCAGGCTCGGCTTCAACTGACACCGGTCTCCCAGGCGTGGAAGCCTCATCCA 156
              || ||||  ||||||||||||||||||||||||||||||||||| |||||||||||||||||| |
Db      67499 AAAGGACTCAGAGGCTCACTTCAACTGCACCACCCAAGGAACCATCATCACCACCATCATGTG 67558

Qy        157 GGCTCTCAGTTGGACATGGTGGTGGCTAAGGGTCAGGCCCATGGAGCCCATATCACCAATGA 216
              |||||||||||||||||||||||||||||| || |||||||||||||| ||||||||| |||||
Db      67559 GACTCTTAACCAAAATGGTGGTTGGCTGAGTCTCAGCAAGGACAGCATCATCACCACAA 67618

Qy        217 CCGCTTCACCTCTCAGAGGTACGACCAGGGGCGAACTTCACCTCGGAGGATGAATCATCCA 276
              |||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||| ||
Db      67619 CCGCTTCACCTATGCCAGTTACACAGCAGCTTCATCTCGGAGTTGATCATCGAATCATCCA 67678

Qy        277 CAATGTGGAGCCCACTGATTCGGGGGAACATCAGATGCAGCCTCCAGAACAGTCGCCCTGCA 336
              |||| ||||| ||||||||||||||||||||||||||||| || |||||||||||||||||| |||
Db      67679 TGATGTGCACCCAGTGACTGCGGGGATCCGTGCAAATGCAGCCTCGCAGAACAGCCCATGGGTT 67738

Qy        337 TGGATCTGCTTACCTTACCGTCCAAGTT 364
              |||||||| ||| |||||||||||||||||
Db      67739 TGGATCTGCCTCCTTCCTCTCAGTGCAAGGT 67766
```

RESULT 10

LOCUS   AX380408   754 bp   DNA   linear   PAT 18-MAR-2002
DEFINITION   Sequence 13 from Patent WO200710.
ACCESSION   AX380408
VERSION   AX380408.1 GI:19575338
KEYWORDS
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
TITLE   B7-like molecules and uses thereof
JOURNAL   Patent: WO 0200710-A 13 03-JAN-2002;
             Amgen, Inc. (US)
FEATURES   Location/Qualifiers
     source   1..754
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS   53..724
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD28222.1"
                     /db_xref="GI:19575339"
                     /translation="MEGSWRDVLAVILAVILAQLTASGSSSYQIIEGPQNVTVLKDSEAHF
                     NCTVTHGWKILMWTLLNQMVVLSLTTQGPIITINNRFTYASYNSTDSFISELIIHDVQPS
                     DSGSVQCSLQNSHGFGSAAFLSVQESTYQNEIRKSANMRTNKADPETKLKSGKENYGYS
                     SDEAKAAQTASLPPKSAEVSLPEKRSSSLPYQELNKHQPGPATHPRVSFDIASPQKVR
                     NVTLV"

BASE COUNT   220 a   191 c   175 g   168 t
ORIGIN

Query Match   15.6%; Score 181.8; DB 6; Length 754;
Best Local Similarity   72.9%; Pred. No. 5.1e-33;
Matches   234; Conservative   0; Mismatches   87; Indels   0; Gaps   0;

Qy   42 CCGGTTCTGGGTTCTGGTAATGAAGTCATAGAAGAGGCCCCCAAAATGCAAGAGTCCTGAAGG 101
          || ||| ||| |||| | ||| ||| || | |||||||||||||||| |||| |||||
Db   108 CAGCTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCGAGAAATGTAACAGTCCTAAAGG 167

Qy   102 GCTCCCAGGGTCGCCTTCAACTGCACCGTCTCCCAGGGCTGGGAAGCTCATCATGTGGGCTC 161
          ||| | || ||| ||||||||||| ||||| ||||| |||||||||||||||| || |
Db   168 ACTCAGAGGGCTCACTTCAACTGCACCGTCACGGCTGGGAAGCTTCATGTGGACTC 227

Qy   162 TCAGTGACATGGTGGTGCTAAGCGTGCAGGCCCATGGAGCCCATCATCACCAATGACCGCT 221
          ||| |||||| | |||| |||||||||||||||| | ||| |||||||||||||| ||
Db   228 TTAACCAAAATGGTGGTGCTCAGTGTCTCCACCACCCAAGGACCCATCATCCAACAACCGCT 287

Qy   222 TCACCCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATCATCCACAATG 281
          ||||| || |||| | | |||||| |||||||||||||| ||| |||||||||||||| |
Db   288 TCACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTGGGAGTTGATCATCCATGATG 347

Qy   282 TGGAGCCCAGTGATTCGGGGAACATCAGATGCAGGGCGCCTCCAGACAGTCGCCTGCATGGAT 341
          || |||||||||| |||||| ||||||| | || ||||||||| ||||| |||||| ||
Db   348 TGGAGCCCAGTGACTCGGGATCCGTGCAAATGCAGCCTGCGAGAAACAGCCCATGGGTTTGGAT 407

Qy   342 CTGCTTACCTTACCGTCCAAG 362
          ||| | ||| || || ||||
Db   408 CTGGCCTTCCTCTCAGTGCAAG 428

RESULT 11

LOCUS   AX380406   895 bp   DNA   linear   PAT 18-MAR-2002
DEFINITION   Sequence 11 from Patent WO200710.
ACCESSION   AX380406
VERSION   AX380406.1 GI:19575336
KEYWORDS
SOURCE   house mouse.
ORGANISM   Mus musculus

REFERENCE   1
AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
TITLE   B7-like molecules and uses thereof
JOURNAL   Patent: WO 0200710-A 11 03-JAN-2002;
             Amgen, Inc. (US)
FEATURES   Location/Qualifiers
     source   1..895
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS   53..865
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD28221.1"
                     /db_xref="GI:19575337"
                     /translation="MEGSWRDVLAVILAVILAQLTASGSSSYQIIEGPQNVTVLKDSEAHF
                     NCTVTHGWKILMWTLLNQMVVLSLTTQGPIITINNRFTYASYNSTDSFISELIIHDVQPS
                     DSGSVQCSLQNSHGFGSAAFLSVQDSIGEEGPALPWAIILLAVAFSLLILIIVLIII
                     FCCCASRREKEESTYQNEIRKSANMRTNKADPETKLKSGKENYGYSSDEAKAAQTAS
                     LPPKSAEVSLPEKRSSSLPYQELNKHQPGPATHPRVSFDIASPQKVRNVTLV"

BASE COUNT   250 a   223 c   212 g   210 t
ORIGIN

Query Match   15.6%; Score 181.8; DB 6; Length 895;
Best Local Similarity   72.9%; Pred. No. 5.2e-33;
Matches   234; Conservative   0; Mismatches   87; Indels   0; Gaps   0;

Qy   42 CCGGTTCTGGGTTCTGGTAATGAAGTCATAGAAGAGGCCCCCAAAATGCAAGAGTCCTGAAGG 101
          || ||| ||| |||| | ||| ||| || | |||||||||||||||| |||| |||||
Db   108 CAGCTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCGAGAAATGTAACAGTCCTAAAGG 167

Qy   102 GCTCCCAGGGTCGCCTTCAACTGCACCGTCTCCCAGGGCTGGGAAGCTCATCATGTGGGCTC 161
          ||| | || ||| ||||||||||| ||||| ||||| |||||||||||||||| || |
Db   168 ACTCAGAGGGCTCACTTCAACTGCACCGTCACGGCTGGGAAGCTTCATGTGGACTC 227

Qy   162 TCAGTGACATGGTGGTGCTAAGCGTGCAGGCCCATGGAGCCCATCATCACCAATGACCGCT 221
          ||| |||||| | |||| |||||||||||||||| | ||| |||||||||||||| ||
Db   228 TTAACCAAAATGGTGGTGCTCAGTGTCTCCACCACCCAAGGACCCATCATCCAACAACCGCT 287

Qy   222 TCACCCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATCATCCACAATG 281
          ||||| || |||| | | |||||| |||||||||||||| ||| |||||||||||||| |
Db   288 TCACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTGGGAGTTGATCATCCATGATG 347

Qy   282 TGGAGCCCAGTGATTCGGGGAACATCAGATGCAGGGCGCCTCCAGACAGTCGCCTGCATGGAT 341
          || |||||||||| |||||| ||||||| | || ||||||||| ||||| |||||| ||
Db   348 TGGAGCCCAGTGACTCGGGATCCGTGCAAATGCAGCCTGCGAGAAACAGCCCATGGGTTTGGAT 407

Qy   342 CTGCTTACCTTACCGTCCAAG 362
          ||| | ||| || || ||||
Db   408 CTGGCCTTCCTCTCAGTGCAAG 428

RESULT 12

LOCUS   BC004806   1556 bp   mRNA   linear   ROD 07-AUG-2002
DEFINITION   Mus musculus, RIKEN cDNA 2010003D20 gene, clone MGC:7960
             IMAGE:3584645, mRNA, complete cds.
ACCESSION   BC004806
VERSION   BC004806.1 GI:13435932
KEYWORDS   MGC.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1556)
AUTHORS   Strausberg,R.
TITLE   Direct Submission
JOURNAL   Submitted (21-MAR-2001) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov

```
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 10 Row: m Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.

FEATURES             Location/Qualifiers
     source          1..1556
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="FVB/N"
                     /clone="MGC:7960 IMAGE:3584645"
                     /tissue_type="Mammary tumor. Metallothionien-TGF alpha
                     model. 10 month old virgin mouse. Taken by biopsy."
                     /clone_lib="NCI_CGAP_Mam1"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS             55..867
                     /codon_start=1
                     /product="RIKEN cDNA 2010003D20 gene"
                     /protein_id="AAH04806.1"
                     /db_xref="GI:13435933"
                     /db_xref="LocusID:72058"
                     /translation="MEGSWRDVLAVIVILAQLTASGSSYQIIEGPQNVTVLKDSEAHF
                     NCTVTHGWKLLMWILNQMVVLSLTTQGPIIITNNRFTVASYNSTDSFISELIIHDVQPS
                     DSGSVQCSLQNSHGPGSAFLSVQDSIGEGPALPTWAIILLAVAFSLLLLILIVLLII
                     LPPKSAEVSLPEKRSSSLPYQELNKHQPGPATHPRVSFDIASPQKVRNVTLV"
BASE COUNT      447 a    364 c    367 g    378 t
ORIGIN
```

```
Query Match     15.6%;  Score 181.8;  DB 10;  Length 1556;
Best Local Similarity  72.9%;  Pred. No. 5.3e-33;
Matches  234;  Conservative  0;  Mismatches  87;  Indels  0;  Gaps  0;

Qy     42 CCGGTTCTCGGGTTCTTGGTAATGAAGTCATAGAAGAGGCCCCCAAAATGCAAGAGTCCTGAAGG 101
          |  ||| ||||||  |||||  |||||||| |||||||||| ||| ||| || |||||| |||||
Db    110 CAGCTTCCGGATCCAGTTATCAGAATCATAGAAGAGGTCCTTCAGAATTGTACAGTCCTAAAGG 169

Qy    102 GCTCCCAGGCTCGCCTTCAACTGACATGCACCGTGTCTCCCAGGGCTGGGAAGGCTCATCAATGTGGGCTC 161
          ||||| ||||||| ||||||||| |  |||||||||||    || | |||||| | ||||| ||||| |||
Db    170 ACTCAGAGGCTCACTTCAACTGCACCGTGTCGACTCAGGCGGTGGAAGCTTCTCATGTGGACTC 229

Qy    162 TCAGTGACATGGGTGGTGCTTAAGGTCAGGCCCATGAGAGCCCATCATCACCAATGACCGGCT 221
          |  ||||| |||| ||||||||| | | |  ||||  |   |||||||||||||  ||| ||||
Db    230 TTAACCAAATGGTGGTGCTGTCTCACCCCAAGGACCCATCATCACCAACACGGCT 289

Qy    222 TCACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGAGATGACATCCACAATG 281
          |||||| | ||| |   |  |||| | | ||  |||| |||||| |||| ||||| ||||
Db    290 TCACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATCCATGATG 349

Qy    282 TGGAGCCCAGTGATTCGGGGAACATCAGATGCCAGCCTCCAGAACAGTCGCCTGCANGGAT 341
          || ||||||||||  ||||| | | ||| ||||||||| ||||||||  | |   ||||||
Db    350 TGCAGCCCAGTGACTCGGGATCGGTGCAAATGCCAGCCTGCAGAACAGCCATGGGTTTGGAT 409

Qy    342 CTGCTTACCTTACCGTCCAAG 362
          ||||  |  || | || ||||
Db    410 CTGCCTTCCTCTCCAGAGTGCAAG 430
```

```
RESULT 13
LOCUS       AC120346              175861 bp    DNA     linear   HTG 09-MAY-2002
DEFINITION  Mus musculus clone RP23-147E11, WORKING DRAFT SEQUENCE, 22 ordered
            pieces.
ACCESSION   AC120346
VERSION     AC120346.3  GI:20514894
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 175861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,R., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   2  (bases 1 to 175861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliey,I., Johnson,R., Jones,C.,
            Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,R., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 175861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliey,I., Johnson,R., Jones,C.,
            Kanat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travis,M., Travis,R., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 9, 2002 this sequence version replaced gi:20503171.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
```

```
      Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
      Project Information
 Center project name: L20921
 Center clone name: 147_E_11
      Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 169879 bases at least Q40
 Consensus quality: 172188 bases at least Q30
 Consensus quality: 173113 bases at least Q20
 Consensus quality: agarose-fp
 Insert size: 176000; agarose-fp
 Insert size: 173761; sum-of-contigs
 Quality coverage: 8.3 in Q20 bases; agarose-fp
 Quality coverage: 8.4 in Q20 bases; sum-of-contigs

 NOTE: This is a 'working draft' sequence. It currently
 consists of 22 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces
 is believed to be correct as given, however the sizes
 of the gaps between them are based on estimates that have
 provided by the submitter.
 This sequence will be replaced
 by the finished sequence as soon as it is available and
 the accession number will be preserved.
```

```
FEATURES             Location/Qualifiers
 source            1..175861
                   /organism="Mus musculus"
                   /db_xref="taxon:10090"
                   /clone="RP23-147E11"
                   /clone_lib="RPCI-23 Female Mouse BAC"
 misc_feature     1..28835
                   /note="assembly_fragment
                   clone_end:SP6
                   vector_side:left"
 misc_feature     28936..29562
                   /note="assembly_fragment"
 misc_feature     29663..30328
                   /note="assembly_fragment"
 misc_feature     30429..32005
                   /note="assembly_fragment"
 misc_feature     32106..33128
                   /note="assembly_fragment"
 misc_feature     33229..34512
                   /note="assembly_fragment"
 misc_feature     34613..36079
                   /note="assembly_fragment"
 misc_feature     36180..37483
                   /note="assembly_fragment"
 misc_feature     37584..39166
                   /note="assembly_fragment"
 misc_feature     39267..41065
                   /note="assembly_fragment"
 misc_feature     41166..43203
                   /note="assembly_fragment"
 misc_feature     43304..47489
                   /note="assembly_fragment"
 misc_feature     47590..50604
                   /note="assembly_fragment"
 misc_feature     50705..56370
                   /note="assembly_fragment"
 misc_feature     56471..64079
                   /note="assembly_fragment"
 misc_feature     64180..72619
                   /note="assembly_fragment"
 misc_feature     72720..81566
                   /note="assembly_fragment"
 misc_feature     81667..96774
                   /note="assembly_fragment"
 misc_feature     96875..119647
                   /note="assembly_fragment"
 misc_feature     119748..147772
                   /note="assembly_fragment"
 misc_feature     147873..174557
                   /note="assembly_fragment"
 misc_feature     174658..175861
                   /note="assembly_fragment
                   clone_end:T7
                   vector_side:right"
 BASE COUNT    48903 a    39540 c   39029 g    46286 t    2103 others
 ORIGIN
```

```
Query Match      14.7%;  Score 171.4;  DB 2;  Length 175861;
Best Local Similarity  72.0%;  Pred. No. 2.1e-30;
Matches  237;  Conservative  0;  Mismatches  91;  Indels  1;  Gaps  1;

Qy          37 CCCACCCGGTTCTGGGTTCTGGTAATGAAGTCATAGAAGGCCCCCAAAATGCAAGAGTCCT  96
                || | |||||  ||||| |||||| |||  | ||||||||||||||||| ||| |||||||
Db      120535 CCTTCTPAGCTTCCGGATCCAGTTATCAGATCAATAGGAAGGTCCTCAGAATGTAACAGTCCT  120476

Qy          97 GAAGGGCTCCCAGGGCTCGCGCTTCAACTGCACCGT-CTCCCAGGGCGTGGGAAGCTCATCATGT  155
                |||| |||| |  ||| || ||||||||||||||| | |||||||||| |||||||| |||||||
Db      120475 AAAGGACTCAGAGGCTCACTTCAACTGCACCGTGGGCTCACGGTGGGAAGCTTCTCAGTGT  120416

Qy         156 GGGCTCTCAGTGACATGGTTGGTTGCTAAGCGTCAGGCCCCATGGAGCCATCATCACCAATG  215
                ||| |||| |||||| |||||| |||| | | || || || |||||||||| ||||||||||
Db      120415 GGACTCTTAACCAAATGGTGGTTGCTGAGTCTTCACCACCCAAGGAGACCCATCATCACCACA  120356

Qy         216 ACCGCTTCACCTCTCAGAGGTACGACCAGGCGCGGGAACTTCACCTCGGAGAGTGATGATCATCC  275
```

```
RESULT 14
AF064857

Db  120355 ACGGCTTCACCTATCCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATCC 120296
           |||||||| ||| |||| | ||| |  |||||||||| ||||  |||||| ||||||||
Qy     276 ACAATGTGGAGCCCAGTGATTCGGGGAACATCAGATGCAGCCTCCAGAACAGTGCGCCTGC 335

Db  120295 ATGATGTGCAGCCCAGTAGCTCGGGGATCCGGTGACTCGGAACTGCCTGCGAGAACACCAC 120236
           |||| | ||||||||||| | ||||||| || ||| | ||||| | |||| || |||||
Qy     336 ATGGAGTCGTTACTTACCGTCACGGTCCAAGTT 364

Db  120235 TTGGATCTGCCTTCCTCTCAGTGCAAGGT 120207
           ||||||||| ||| ||||| ||| |||||
Qy     336                                364
```

```
LOCUS       AF064857             199665 bp    DNA     linear   PRI 23-JUN-1998
DEFINITION  Homo sapiens BAC derived from chromosome 21q22.3, complete
            sequence, containing PEP19 (PCP4) gene.
ACCESSION   AF064857
VERSION     AF064857.1  GI:3171149
KEYWORDS    HTG.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 199665)
  AUTHORS   Taudien,S., Nordsiek,G., Korenberg,J., Drescher,B., Weber,J.,
            Schattevoy,R. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAY-1998) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1..199665
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
     repeat_region   complement(1..383)
                     /evidence=not_experimental
                     /rpt_family="L1"
     repeat_region   747..814
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(1345..1500)
                     /evidence=not_experimental
                     /rpt_family="L1PA4"
     repeat_region   1499..1588
                     /evidence=not_experimental
                     /rpt_family="L1PA2"
     repeat_region   complement(1584..2179)
                     /evidence=not_experimental
                     /rpt_family="L1PA2"
     repeat_region   2174..3405
                     /evidence=not_experimental
                     /rpt_family="L1PA2"
     repeat_region   3388..3454
                     /evidence=not_experimental
                     /rpt_family="L1P"
     repeat_region   3968..4086
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(6245..6668)
                     /evidence=not_experimental
                     /rpt_family="L1ME3A"
     repeat_region   6671..6969
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   complement(7057..7376)
                     /evidence=not_experimental
                     /rpt_family="MER2"
     repeat_region   complement(7387..7491)
                     /evidence=not_experimental
                     /rpt_family="L1ME3A"
     repeat_region   7781..8020
```

```
                     /evidence=not_experimental
                     /rpt_family="MER58A"
     repeat_region   8650..8823
                     /evidence=not_experimental
                     /rpt_family="L1PA5"
     repeat_region   complement(10456..10720)
                     /evidence=not_experimental
                     /rpt_family="MER5A"
     repeat_region   complement(12514..13814)
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   complement(13823..14116)
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(15662..15785)
                     /evidence=not_experimental
                     /rpt_family="AluJo"
     repeat_region   complement(18994..19347)
                     /evidence=not_experimental
                     /rpt_family="MER3"
     repeat_region   19378..19470
                     /note="homology = 81.70%, score = 28, counts = 3"
                     /evidence=not_experimental
                     /rpt_family="THE1B"
     repeat_region   21213..21500
                     /evidence=not_experimental
                     /rpt_family="MER3"
     repeat_region   complement(22371..20556)
                     /rpt_unit=agtatctatctgattgggcagtgctcatac
                     /rpt_type=tandem
     repeat_region   22291..22567
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(22583..23201)
                     /evidence=not_experimental
                     /rpt_family="MER39b"
     repeat_region   23202..23471
                     /evidence=not_experimental
                     /rpt_family="L1PA2"
     repeat_region   24223..24454
                     /evidence=not_experimental
                     /rpt_family="MER39b"
     repeat_region   24455..24889
                     /evidence=not_experimental
                     /rpt_family="MSTB-internal"
     repeat_region   complement(25434..25767)
                     /evidence=not_experimental
                     /rpt_family="MSTB"
     repeat_region   25993..26294
                     /evidence=not_experimental
                     /rpt_family="MER58B"
     repeat_region   27222..27283
                     /evidence=not_experimental
                     /rpt_family="AluSq"
     repeat_region   complement(28932..29335)
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(29462..30001)
                     /evidence=not_experimental
                     /rpt_family="MSTA"
     repeat_region   complement(30308..30676)
                     /evidence=not_experimental
                     /rpt_family="MER34"
     repeat_region   31012..31364
                     /evidence=not_experimental
                     /rpt_family="MLT1B"
     repeat_region   complement(32507..32738)
                     /evidence=not_experimental
                     /rpt_family="MLT1A2"
     repeat_region   complement(32766..32980)
                     /evidence=not_experimental
                     /rpt_family="MSTD"
```

```
repeat_region    33468..33678
                 /evidence=not_experimental
                 /rpt_family="THE1B"
repeat_region    34563..34620
                 /evidence=not_experimental
                 /rpt_family="MLT2CB"
                 /rpt_type=tandem
                 /rpt_unit=tttccagaaaagaataacgtacttaggca
                 /note="homology = 93.10%, score = 21, counts = 2"
repeat_region    complement(35785..35824)
                 /evidence=not_experimental
                 /rpt_family="MLT2CB"
repeat_region    complement(35924..36345)
                 /evidence=not_experimental
                 /rpt_family="MLT2CB"
repeat_region    complement(36403..36475)
                 /evidence=not_experimental
                 /rpt_family="MLT2CB"
repeat_region    complement(36717..36828)
                 /evidence=not_experimental
                 /rpt_family="MLT2CDS"
repeat_region    complement(37064..37364)
                 /evidence=not_experimental
                 /rpt_family="MER67C"
repeat_region    complement(37440..37501)
                 /evidence=not_experimental
                 /rpt_family="AluY"
                 /rpt_type=tandem
                 /rpt_unit=aaaacacttcacttgtcttcaaaactgctc
                 /note="homology = 98.40%, score = 29, counts = 2"
repeat_region    37518..38158
                 /evidence=not_experimental
                 /rpt_family="L2"
repeat_region    38477..39376
                 /evidence=not_experimental
                 /rpt_family="L2"
repeat_region    39552..39861
                 /evidence=not_experimental
                 /rpt_family="L2"
repeat_region    39884..40022
                 /evidence=not_experimental
                 /rpt_family="MLT1D"
repeat_region    complement(40243..40542)
                 /evidence=not_experimental
                 /rpt_family="MLT1D"
repeat_region    4054..4180
                 /evidence=not_experimental
                 /rpt_family="AluSx"
repeat_region    complement(42066..42114)
                 /evidence=not_experimental
                 /rpt_family="L2"
repeat_region    42138..42317
                 /evidence=not_experimental
                 /rpt_family="MIR"
repeat_region    complement(42332..42710)
                 /evidence=not_experimental
                 /rpt_family="L2"
repeat_region    complement(42758..46751)
                 /evidence=not_experimental
                 /rpt_family="LIPB1"
repeat_region    complement(46792..47084)
                 /evidence=not_experimental
                 /rpt_family="LIPB1"
repeat_region    47189..47992
                 /evidence=not_experimental
                 /rpt_family="L1MA"
repeat_region    complement(48002..48132)
                 /evidence=not_experimental
                 /rpt_family="L2"
repeat_region    48905..49310
                 /evidence=not_experimental
                 /rpt_family="MIR"
```

```
Qy   1069  CCAGCGGGCGGTCGATCANCGTCCACCCAGGCGCAGGCGCAAGTCATCCACACAGGCTTCTTTTTAATCT 1128
           |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db  17555  CCAGCGGGCGGTCGATCAACGTCCACCCAGGCGCAGGCGCAAGTCATCCACACAGGCTTCTTTTTAATCT 17614

Qy   1129  GGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTAG 1168
           ||||||||||||||||||||||||||||||||||||||
Db  17615  GGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTAG 17654
```

```
RESULT 15
AC020851/c
LOCUS       AC020851              196900 bp    DNA     linear   HTG 15-JUL-2000
DEFINITION  Mus musculus clone RP21-467L12, WORKING DRAFT SEQUENCE, 33
            unordered pieces.
ACCESSION   AC020851
VERSION     AC020851.2  GI:9211212
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 196900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 15, 2000 this sequence version replaced gi:6686457.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 1437240
            Center Project Name:
            Center clone name: RPCI-21_467L12
            -----
            Summary Statistics
            Consensus quality: 164618 bases at least Q40
            Consensus quality: 180887 bases at least Q30
            Consensus quality: 183777 bases at least Q20
            Estimated insert size: 147000; pulse field gel estimation
            Estimated insert size: 193700; sum-of-contigs estimation
            Quality coverage: 5.67 in Q20 bases; pulse field gel estimation
            Quality coverage: 4.3 in Q20 bases; sum-of-contigs estimation.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 33 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *        1    1541: contig of 1541 bp in length
          *     1542    1641: gap of unknown length
          *     1642    3492: contig of 1851 bp in length
          *     3493    3592: gap of unknown length
          *     3593    4738: contig of 1146 bp in length
          *     4739    4838: gap of unknown length
          *     4839    7248: contig of 2410 bp in length
          *     7249    7348: gap of unknown length
          *     7349    9178: contig of 1830 bp in length
          *     9179    9278: gap of unknown length
          *     9279   11493: contig of 2215 bp in length
          *    11494   11593: gap of unknown length
          *    11594   13607: contig of 2014 bp in length
          *    13608   13707: gap of unknown length
          *    13708   15946: contig of 2239 bp in length
          *    15947   16046: gap of unknown length
          *    16047   18392: contig of 2346 bp in length
```

```
Query Match       8.6%;  Score 100;  DB 9;  Length 199665;
Best Local Similarity 100.0%;  Pred. No. 3.3e-13;
Matches 100;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;
```

Search completed: April 29, 2003, 14:30:21
Job time : 4416.94 secs

```
FEATURES             Location/Qualifiers
     source          1..196900
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP21-467L12"
                     /clone_lib="RPCI mouse PAC library 21"
BASE COUNT     51456 a   43823 c   44608 g   53666 t   3347 others
ORIGIN

*    18393
*    18493    18492: gap of unknown length
*    21232    21231: contig of 2739 bp in length
*    21332    21331: gap of unknown length
*    24663    24662: contig of 3331 bp in length
*    24763    24762: gap of unknown length
*    27141    27140: contig of 2378 bp in length
*    27241    27240: gap of unknown length
*    29951    29950: contig of 2710 bp in length
*    30051    30050: gap of unknown length
*    32246    32245: contig of 2195 bp in length
*    32346    32345: gap of unknown length
*    35689    35688: contig of 3343 bp in length
*    35789    35788: gap of unknown length
*    40388    40387: contig of 4599 bp in length
*    40488    40487: gap of unknown length
*    43581    43580: contig of 3093 bp in length
*    43681    43680: gap of unknown length
*    46438    46437: contig of 2757 bp in length
*    46538    46537: gap of unknown length
*    50185    50184: contig of 3647 bp in length
*    50285    50284: gap of unknown length
*    54573    54572: contig of 4288 bp in length
*    54673    54672: gap of unknown length
*    60228    60227: contig of 5555 bp in length
*    60328    60327: gap of unknown length
*    68025    68024: contig of 7697 bp in length
*    68125    68124: gap of unknown length
*    78611    78610: contig of 10486 bp in length
*    78711    78710: gap of unknown length
*    88252    88251: contig of 9441 bp in length
*    88352    88351: gap of unknown length
*    96258    96257: contig of 8006 bp in length
*    96358    96357: gap of unknown length
*   106228   106227: contig of 9870 bp in length
*   106328   106327: gap of unknown length
*   115896   115895: contig of 9468 bp in length
*   115996   115995: gap of unknown length
*   126364   126363: contig of 10468 bp in length
*   126464   126463: gap of unknown length
*   137190   137189: contig of 10726 bp in length
*   137290   137289: gap of unknown length
*   146379   146378: contig of 9089 bp in length
*   146479   146478: gap of unknown length
*   161687   161686: contig of 15208 bp in length
*   161787   161786: gap of unknown length
*   176150   176149: contig of 14363 bp in length
*   176250   176249: gap of unknown length
*   196900   196900: contig of 20651 bp in length.
```

```
Query Match     4.9%;   Score 57.4;   DB 2;   Length 196900;
Best Local Similarity   66.7%;
Matches   82;   Conservative   0;   Mismatches  41;   Indels   0;   Gaps   0;
                                     Pred. No. 0.0058;

Qy    1045 AAACAGTAGCTGTGGCCCTCCTCACCACGCGGCGTGATCAACGTCACCCAGGCCAGCAAG 1104
           | | ||| |  ||| | | |||||||||||||| |||||  ||||| |||||||||
Db  119051 ACACTGACCCTCTACCTTTGTTTACCAGGAACTCAATAAACATCAGCCCGGTCCAGCAAC 118992

Qy    1105 TCATCCACAGGCTTCTTTTAATCTGGCCCAGTCCTGAGAAGGTCAGTAATACAACGTAGT 1164
           ||||||||| ||||||||||||| ||||||||||| |||||| |||| ||| |||||||
Db  118991 TCATCCACGGGTTTCCTTTGACATCGCCAGTCCTCAGAGGTCAGAAATGTGACTTTAGT 118932

Qy    1165 ATA 1167
           |||
Db  118931 GTA 118929
```

OM protein - protein search, using sw model

```
                                        GenCore version 5.1.4_p5_4578
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:             April 28, 2003, 16:24:30 ; Search time 10.2617 seconds
                                              (without alignments)
                                              1543.990 Million cell updates/sec

Title:
Perfect score:      2059
Sequence:           1 MGLVIFLHGSGSGGNEVIEGP..........HPQASFNLASPEKVSNTTVV 382

Scoring table:      BLOSUM62
                    Gapop 10.0 , Gapext 0.5
Searched:           112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters:        112892

Minimum DB seq length:    0
Maximum DB seq length:    200000000000

Post-processing:    Minimum Match 0%
                    Maximum Match 100%
                    Listing first 45 summaries

Database :          SwissProt_40:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB %	ID	Description
1	147.5	7.2	404	1	Q15109;	RAGE_HUMAN homo sapien
2	131	6.4	337	1	Q98892	G55A_CHICK gallus gall
3	129.5	6.3	344	1	Q62218	NTRI_RAT rattus norv
4	128	6.3	353	1	Q90773	CEPU_CHICK gallus gall
5	128	6.2	1447	1	Q15746	DCC_HUMAN homo sapien
6	128	6.2	1914	1	Q15746	KMLS_HUMAN homo sapien
7	127	6.2	1447	1	P70211	DCC_MOUSE mus musculu
8	124	6.0	1070	1	Q13308	PTK7_HUMAN homo sapien
9	120	6.0	345	1	Q14992	OPCM_HUMAN homo sapien
10	119	5.8	345	1	P11834	OPCM_BOVIN bos taurus
11	118	5.7	345	1	P32736	OPCM_RAT rattus norv
12	117.5	5.7	345	1	Q90610	NEO1_CHICK gallus gall
13	116	5.6	4393	1	P98160	PGBM_HUMAN homo sapien
14	114.5	5.6	1461	1	Q92859	NEO1_HUMAN homo sapien
15	113.5	5.5	416	1	Q28173	RAGE_BOVIN bos taurus
16	113	5.5	1377	1	Q97603	NEO1_RAT rattus norv
17	112	5.4	764	1	Q08180	ICCR_DROME drosophila
18	110.5	5.4	524	1	Q62556	BUTY_MOUSE mus musculu
19	110.5	5.4	620	1	Q92154	SMP_COTJA coturnix co
20	109.5	5.3	249	1	Q03751	CSP_DROME drosophila
21	107	5.2	365	1	P78310	CXAR_HUMAN homo sapien
22	107	5.2	3707	1	Q05793	PGBM_MOUSE mus musculu
23	106.5	5.2	333	1	P15364	AMAL_DROME drosophila
24	106	5.1	319	1	Q99795	A33_HUMAN homo sapien
25	106	5.1	1092	1	P36635	NCA2_XENLA xenopus lae
26	104.5	5.1	862	1	Q62786	CD22_MOUSE mus musculu
27	104.5	5.1	879	1	P35329	FPR2_RAT rattus norv
28	103.5	5.0	338	1	Q98919	LAMP_CHICK gallus gall
29	103.5	5.0	890	1	Q06418	TYO3_HUMAN homo sapien
30	102.5	5.0	344	1	P08921	CD2_RAT rattus norv
31	102.5	5.0	348	1	Q9z018	KILO_RAT rattus norv
32	102.5	5.0	365	1	P97792	CXAR_MOUSE mus musculu
33	102.5	5.0	2499	1	P08169	MPRI_BOVIN bos taurus
34	102	5.0	880	1		TYO3_RAT rattus norv
35	100.5	4.9	873	1	P55146	FAS2_DROME drosophila
36	100	4.9	1241	1	P34082	NPHN_HUMAN homo sapien
37	99.5	4.8	261	1	O60500	KLK8_RAT rattus norv
38	99.5	4.8	1260	1	P36374	CAML_MOUSE mus musculu
39	99.5	4.8	1906	1	P11627	KMLS_CHICK gallus gall
40	99.5	4.8	2481	1	P11799	UN52_CAEEL caenorhabdi
41	99	4.8	338	1	O05561	UN52_CAEEL caenorhabdi
42	99	4.8	338	1	O05561	LAMP_HUMAN homo sapien
43	99	4.8	359	1	Q13449	LAMP_RAT rattus norv
44	99	4.8	830	1	O62813	LACH_DROME drosophila
45	99	4.8	880	1	Q24372	SREC_HUMAN homo sapien
				1	Q14162	TYO3_HUMAN homo sapien
				1	P55144	TYO3_MOUSE mus musculu

ALIGNMENTS

RESULT 1
```
ID   RAGE_HUMAN     STANDARD;      PRT;    404 AA.
AC   Q15109; Q15279; Q9Y3R3; Q9H2X7;
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=95137587; PubMed=7835890;
RA   Sugaya K., Fukagawa T., Matsumoto K., Mita K., Takahashi E., Ando A.,
RA   Inoko H., Ikemura T.;
RT   "Three genes in the human MHC class III region near the junction with
RT   the class II: gene for receptor of advanced glycosylation end
RT   products, PBX2 homeobox gene and a notch homolog, human counterpart
RT   of mouse mammary tumor gene int-3.";
RL   Genomics 23:408-419(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=92340547; PubMed=1378843;
RA   Neeper M., Schmidt A.M., Brett J., Yan S.D., Wang F., Pan Y.C.,
RA   Elliston K., Stern D., Shaw A.;
RT   "Cloning and expression of a cell surface receptor for advanced
RT   glycosylation end products of proteins.";
RL   J. Biol. Chem. 267:14998-15004(1992).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA   Banta A., Spies T., Hood L.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT ARG-100.
RA   Abedin M.J., Yonekura H., Migita H., Karasawa J., Yamamoto Y.,
RA   Yamamoto H.;
RT   "Molecular heterogeneity of the receptor for advanced glycation
RT   endproducts.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Malherbe P., Richards J., Gaillard H., Thompson A., Diener C.,
RA   Schuler A., Huber G.;
RT   "cDNA cloning of a novel secreted isoform of the human Receptor for
RT   Advanced Glycation End products (RAGE) and characterization of cells
RT   co-expressing cell-surface scavenger receptors and Swedish mutant
RT   amyloid precursor protein.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
```

DE   (Receptor for advanced glycosylation end products).
DE   Advanced glycosylation end product-specific receptor precursor
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Created)

GN   AGER OR RAGE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]

```
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 1-12 FROM N.A.
RA   Hudson B.I., Futers T.S.;
RT   "Novel polymorphisms in the receptor for advanced glycation
RT   end-products (RAGE) gene.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MEDIATES INTERACTIONS OF ADVANCED GLYCOSYLATION END
CC       PRODUCTS (AGE). THESE ARE NONENZYMATICALLY GLYCOSYLATED PROTEINS
CC       WHICH ACCUMULATE IN VASCULAR TISSUE IN AGING AND AT AN ACCELERATED
CC       RATE IN DIABETES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1).
CC       Secreted (isoform 2).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2/RAGESEC;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ENDOTHELIAL CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---------------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M91211; AAA03574.1; -.
DR   EMBL; D28770; BAA05958.1; -.
DR   EMBL; U89336; AAB47491.1; -.
DR   EMBL; AB036432; BAA89369.1; -.
DR   EMBL; AJ133822; CABA43108.1; -.
DR   EMBL; BC020669; AAH20669.1; -.
DR   EMBL; AF208289; AAG35728.1; -.
DR   Genew; HGNC:320; AGER.
DR   MIM; 600214; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Immunoglobulin domain; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Alternative splicing; Polymorphism.
FT   SIGNAL        1     22
FT   CHAIN        23    404       ADVANCED GLYCOSYLATION END PRODUCT-
FT                                SPECIFIC RECEPTOR.
FT   DOMAIN       23    342       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    343    363       POTENTIAL.
FT   DOMAIN      364    404       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       31    106       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      137    215       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      252    308       IG-LIKE C2-TYPE DOMAIN 2.
FT   DISULFID     38     99       POTENTIAL.
FT   DISULFID    144    208       POTENTIAL.
FT   DISULFID    259    301       POTENTIAL.
FT   CARBOHYD     25     25       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     81     81       N-LINKED (GLCNAC...) (POTENTIAL).
FT   POLY-GLU    380    384
FT   VARSPLIC     54     67       MISSING (IN ISOFORM 2).
FT   VARSPLIC    275    404       GVPLPLPPSPVLILPEIGPQDGGTYSCVATHSSHQPOESRA
FT                                VSISIIEPGEEGPTAGSVGGSGLGTLALALGLGLGTNAL
FT                                LIGVILWQRQRGEERKAPENQEEEEERAELNQSEEPEAG
FT                                ESSTGGP -> VSDLERGAGRTRRGGANCRLCGRIRAGNS
FT                                PGPGDGPGPGDSRPAHWGHLVAKAATPRGEEGRPKPGGRG
FT                                GACRIESVGGTI (IN ISOFORM 2).
FT   VARIANT     100    100       Q -> R.
FT                                /FTId=VAR_011338.
FT   CONFLICT      1      1       M -> G (IN REF. 1).
```

```
SQ   SEQUENCE   404 AA;  42802 MW;  0D584C436C30CCE7 CRC64;

     Query Match              7.2%;  Score 147.5;  DB 1;  Length 404;
     Best Local Similarity   22.6%;  Pred. No. 0.00012;
     Matches      85;  Conservative    40;  Mismatches 114;  Indels 137;  Gaps 18;

Qy    15  EVIEGPQNATVLKGSQARNCTVSQG---WKLIMWALSDMVVLSVRPMEPIITNDRFTS-   70
          :|:|                 |||       |: |: |:  ::|   |::||
Db   125  EIVDSASELTA--GVPNKYGTCVSEGSYPAGTLSWHLDG------KPLVPNEKGVSV    173

Qy    71  ---QRYDQGGNFT--SEMIIHNVEPSDSGNIR---CSLQNSRLHGSAYLIVQVMGELF   120
             |: |    | :   :|| :| |:|   |||     :| | :       :
Db   174  KEQTRRHPETGLFTLQSELM---VTPARGGDRPTTFSCSFSPGLPRHRALRTAPIQPRVW  230

Qy   121  IP------SVNLVVAENEP-----------CEVTLPSHWTRLPDISWELGLLVSHSSY   162
          ||       |||: |:|            |||  | | |  | ||  || ||  |
Db   231  EPVPLEEVQLVV--EPEGGAVAPGGTVTLTCEVPAQPS-----PQIHWMKD------  274

Qy   163  YFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSAFVNLIVIRCPQDTGGINI  222
           ||  |    : |:|  |: :  ||| |||  |  |::     :|:|      |  ||:
Db   275  -GVPLPPSPVLILPEIGPQDGGTYSCVATHSSHQPOESRAVSIIIE-PGEEG-----  327

Qy   223  PGVLSSLPSLGFSLPTWGKVGLGLAGTMLLI---------PTCTLITIRCCCCRRCCGCNCC  275
           |: | | |  :| | | |||| ||  :||:          | |: :|||| ||| ||| ||
Db   328  -PTAGSVGGSLGTLALALGLGLGTNALLIGVILMQRR-----------  366

Qy   276  CRCCFCCRRKRGFRIQPOKKSEKEKT--NKTETETESGNENSGYNSDEQKTTDTASLPPKS  333
           | | | |::| | :|:|::|::|::|    |::|:  |  :|:  |::|||:||:|||||:
Db   367  -------QRRGEERKAPENQEEEEERAELN-----  389

Qy   334  CESSSDEQRNSSCGPP   349
           | || | :| || ||
Db   390  -QSEEPEAGESSTGGP  404

RESULT 2
ID   G55A_CHICK     STANDARD;      PRT;   337 AA.
AC   Q98892;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neurite inhibitor GP55-A precursor (OBCAM protein gamma isoform).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE OF 64-337 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=97157768; PubMed=9004047;
RA   Wilson D.J.A., Kim D.S., Clarke G.A., Marshall-Clarke S., Moss D.J.;
RT   "A family of glycoproteins (GP55), which inhibit neurite outgrowth,
RT   are members of the Ig superfamily and are related to OBCAM from
RT   chick: structural diversity of IgLON family proteins.";
RL   J. Cell Sci. 109:3129-3138(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Lodge A.P., Kim D.S., Howard M.R., McNamee C.J., Smith N., Moss D.J.;
RT   "Cloning of CEPU-s, a secreted isoform of CEPU-1, and OBCAM cDNAs from
RT   chick.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INHIBITS NEURITE OUTGROWTH.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: RESTRICTED TO THE NERVOUS SYSTEM.
CC   -!- DEVELOPMENTAL STAGE: INCREASES DURING DEVELOPMENT FROM VERY LOW
CC       LEVELS AT EMBRYONIC DAY 10 AND IS MOST ABUNDANT AFTER HATCHING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- ...
```

CC   This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Y08170; CAB41420.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL   1   20
FT   CHAIN   21   317   BY SIMILARITY.
FT   PROPEP   318   337   NEURITE INHIBITOR GP55-A (POTENTIAL).
FT   DOMAIN   43   115   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN   143   201   IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN   229   295   IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID   50   108   POTENTIAL.
FT   DISULFID   150   194   POTENTIAL.
FT   DISULFID   236   288   POTENTIAL.
FT   CARBOHYD   133   133   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   277   277   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   285   285   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   298   298   N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   337 AA;   36887 MW;   BAE717551856651E CRC64;

```
Query Match               6.4%;
Best Local Similarity     25.8%;   Score 131;   DB 1;   Length 337;
Matches          65;   Conservative   33;   Mismatches   96;   Indels   58;   Gaps   14;

Qy   22  NATVLKGSQARFNCTVSQGWKLLMWALSDMVVL-----------------SVRPMEPIITNDRFTSQRYD  74
Db   37  NVTVRQGESATLRCTVDDRVRVAW-LNRSTILYAGNDKWSIDNRVVILSN---TKTQY-  91

Qy   75  QGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSRLHGSAVLTVQVMGELFIPSVNLVV  128
Db   92  -SIKIHNVDVYDEGPYTCSVQTDNHPKTSRVH-LIVQVPPQIVNISSDITV  140

Qy   129 AENEPCEVTCL-----PSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPOS  184
Db   141 NEGSSVTLMCLAFGRPE----PTVTWR--HLSGKQGFVSEDEYLE----ITGITREQ  188

Qy   185 NGTLTCVATWKSLKARKSATVNLTV-------IRCPODTGGGINIPGVL----SSLPSLGFS  235
Db   189 SGEYECSAV-NDVAVPDVRKVKVTVNYPPYISNAKNTGASVGQKGILQCEASAVPVAEFQ  247

Qy   236 LPTWGKVGLGLA  247
Db   248 ---WFKEDTRLA  256
```

RESULT 3
ID   NTRI_RAT   STANDARD;   PRT;   344 AA.
AC   Q62718;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Neurotrimin precursor (GP65).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 217-229.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=95198094; PubMed=7891157;
RA   Struyk A.F., Canoll P.D., Wolfgang M.J., Rosen C.L., D'Eustachio P.,
RA   Salzer J.L.;
RT   "Cloning of neurotrimin defines a new subfamily of differentially
RT   expressed neural cell adhesion molecules.";
RL   J. Neurosci. 15:2141-2156(1995).
-!- FUNCTION: NEURAL CELL ADHESION MOLECULE.
-!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
-!- TISSUE SPECIFICITY: CENTRAL NERVOUS SYSTEM.
-!- DEVELOPMENTAL STAGE: EXPRESSED AT HIGH LEVELS IN SEVERAL
    DEVELOPING PROJECTION SYSTEMS: IN NEURONS OF THE THALAMUS,
    SUBPLATE, AND LOWER CORTICAL LAMINAE IN THE FOREBRAIN AND IN THE
    PONTINE NUCLEUS, CEREBELLAR GRANULE CELLS, AND PURKINJE CELLS IN
    THE HINDBRAIN.
-!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
    SUBFAMILY.
-!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; U16845; AAA67445.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL   1   31
FT   CHAIN   32   321
FT   PROPEP   322   344
FT   DOMAIN   50   122   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN   150   208   IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN   236   302   IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID   57   115   POTENTIAL.
FT   DISULFID   157   201   POTENTIAL.
FT   DISULFID   243   295   POTENTIAL.
FT   CARBOHYD   44   44   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   70   70   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   152   152   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   216   216   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   284   284   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   292   292   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   305   305   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   321   321   N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID   321   321   GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   344 AA;   37998 MW;   CBB3BE53B33B224 CRC64;

```
Query Match               6.3%;
Best Local Similarity     24.2%;   Score 129.5;   DB 1;   Length 344;
Matches          54;   Conservative   43;   Mismatches   104;   Indels   53;   Gaps   14;

Qy   4   VIFLHGSG------SGNEVI-EGPONATVLKGSQARFNCTVSQGWKLIMWALSDMVVLS---  55
Db   21  LIFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDNRVTRVAWLNRSTILYAGND  80

Qy   56  ---VRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSRLH  106
Db   81  KWCLDPRVVLLSN----TQTQY-----SIEIQNVDVYDEGPYTCSVQTDNHPKTSRVH  129

Qy   107 GSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVP  166
Db   130 ---LIVQVSPKIVEISSDISINEGNNISLTCIATGRPE--PTVTWR---HISPKAVGFVS  181

Qy   167 EPSDLQSAVSILALFPQSNGTLTCVATWKSLKARKSATVNLTVIRCP-----ODTGGGIN  221
Db   182 EDEYLE---IQGITREQSGEYECSAS-NDVAAPVVRRVNVTVNYPPYISEAKGTGVPVG  236
```

```
Qy    222  IPGVL----SSLPSLGFSLPTWGK  241
               : :|| |  |||  |  |  |
Db    237  QKGTLQCEASAVPSAEFQ---WFK  257

RESULT 4
CPU_CHICK
ID    CPU_CHICK           STANDARD;    PRT;    353 AA.
AC    Q90773;
DT    01-NOV-1997 (Rel. 35, Created)
DT    01-NOV-1997 (Rel. 35, Last sequence update)
DT    15-JUN-2002 (Rel. 41, Last annotation update)
DE    CPU-1 protein precursor.
OS    Gallus gallus (Chicken).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC    Gallus.
OX    NCBI_TaxID=9031;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=Brain;
RX    MEDLINE=96370549; PubMed=8774445;
RA    Spaltmann F., Bruemmendorf T.;
RT    "CPU-1, a novel immunoglobulin superfamily molecule, is expressed by
RT    developing cerebellar Purkinje cells.";
RL    J. Neurosci. 16:1770-1779(1996).
CC    -!- FUNCTION: IT MAY BE A CELLULAR ADDRESS MOLECULE SPECIFIC TO
CC        PURKINJE CELLS. IT MAY REPRESENT A RECEPTOR OR A SUBUNIT OF A
CC        RECEPTOR COMPLEX.
CC    -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC    -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A MAJOR FORM AND A MINOR FORM
CC        (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC    -!- TISSUE SPECIFICITY: FOUND ON THE DENDRITES, SOMATA AND AXONS OF
CC        DEVELOPING PURKINJE CELLS. UNDETECTABLE ON OTHER NEURONS LIKE
CC        GOLGI OR GRANULE CELLS.
CC    -!- DEVELOPMENTAL STAGE: EXPRESSED BY DEVELOPING CEREBELLAR PURKINJE
CC        CELLS. EXPRESSION COINCIDES WITH THE GROWTH OF THE DENDRITIC TREE,
CC        AFTER PURKINJE CELLS HAVE FINISHED THEIR MIGRATION FROM THE
CC        VENTRICULAR ZONE (FROM E15 UNTIL E21). EXPRESSED IN THE ADULT.
CC    -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC        SUBFAMILY.
CC    -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; Z72497; CAA96578.1; -.
DR    InterPro; IPR003006; Ig_MHC.
DR    InterPro; IPR003598; Ig_c2.
DR    InterPro; IPR003600; Ig_like.
DR    Pfam; PF00047; ig; 3.
DR    SMART; SM00410; IG_like; 1.
DR    SMART; SM00408; IGc2; 2.
KW    Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW    Repeat; Signal; Alternative splicing.
FT    SIGNAL         1     28       POTENTIAL.
FT    CHAIN         29    330       CPU-1 PROTEIN.
FT    PROPEP       331    353       REMOVED IN MATURE FORM (POTENTIAL).
FT    DOMAIN        48    120       IG-LIKE C2-TYPE DOMAIN 1.
FT    DOMAIN       148    206       IG-LIKE C2-TYPE DOMAIN 2.
FT    DOMAIN       234    300       IG-LIKE C2-TYPE DOMAIN 3.
FT    DISULFID      55    113       POTENTIAL.
FT    DISULFID     155    199       POTENTIAL.
FT    DISULFID     241    293       POTENTIAL.
FT    CARBOHYD      42     42       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD      68     68       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     150    150       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     282    282       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     290    290       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     303    303       N-LINKED (GLCNAC...) (POTENTIAL).
FT    LIPID        330    330       GPI-ANCHOR (POTENTIAL).
FT    VARSPLIC     310    320       MISSING (IN MAJOR ISOFORM).
SQ    SEQUENCE   353 AA;   38736 MW;  2550C48591EBBBA6 CRC64;

RESULT 5
DCC_HUMAN
ID    DCC_HUMAN           STANDARD;    PRT;    1447 AA.
AC    P43146;
DT    01-NOV-1995 (Rel. 32, Created)
DT    01-NOV-1995 (Rel. 32, Last sequence update)
DT    15-JUN-2002 (Rel. 41, Last annotation update)
DE    Tumor suppressor protein DCC precursor (Colorectal cancer suppressor).
GN    DCC.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=95011532; PubMed=7926722;
RA    Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA    Vogelstein B.;
RT    "The DCC gene product in cellular differentiation and colorectal
RT    tumorigenesis.";
RL    Genes Dev. 8:1174-1183(1994).
RN    [2]
RP    SEQUENCE OF 1-750 FROM N.A.
RX    MEDLINE=90100559; PubMed=2294591;
RA    Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA    Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA    Vogelstein B.;
RT    "Identification of a chromosome 18q gene that is altered in
RT    colorectal cancers.";
RL    Science 247:49-56(1990).
RN    [3]
RP    SEQUENCE OF 107-472 FROM N.A. (SCRAMBLED EXONS).
RX    MEDLINE=91121517; PubMed=1991322;
RA    Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,
RA    Oliner J.D., Kinzler K.W., Vogelstein B.;
RT    "Scrambled exons.";
RL    Cell 64:607-613(1991).
RN    [4]
RP    GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375.
RX    MEDLINE=94245241; PubMed=8188295;
RA    Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
```

Query Match 63; Conservative 36; Mismatches 90; Indels 56; Gaps 13;
Best Local Similarity 25.7%;
Matches 63;

Score 128; DB 1; Length 353;
Pred. No. 0.0038;

```
Qy     22  NATVLKGSQARENCTVSQ-----GW---KLIMWALSDMVVLSVRPMEPIITNDRFTSQRY   73
               |:|  |  :|  :   |:       | :|| ||  :|  |  |:::   |:  |::
Db     42  NVTVRQGESATLRCSVDNRVTRVAWLNRSSLIYAGNDKWCLDPRVVLLANTKTQYSIQ--   99

Qy     74  DQGGNFTSEMIIHNVEPSDSGNIRCSLQ-----------NSRLHGSAYLTVQVMGELFIPSVNLV   127
               |:|| | |  |||| :| | ||  ||            |||  |  |||    |:|  |
Db    100  --IHDVDYDEGGPYTCSVQTDNHPKTSRVH----LIVQVSPKITETSSDIS   144

Qy    128  VAENEPECVTCLPSHWTRLPD--ISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSN   185
               |:    | || |  | :|  ::  | || |  ||   : |||  | ::| : | |
Db    145  INEGGNVSLTCIA---TGRDPTITWR--HISPKAVGFISEDEYLE---ITGIIREQS   194

Qy    186  GTLTCVATWKSLKARKSATVNLITV-----IRCPDTGGGINIPGVL----SSLPSLGFSL   236
               |  |  :| |||   :| || ||    |  | ||   | ||    |||| | ||
Db    195  GEYEGCSAS-NDVAAPVVQRVKVIVNYPPYISDAKSTGVPVGQKGILMCEASAVPSADFQ-   252

Qy    237  PTWGK  241
           || |
Db    253  --WFK  255
```

```
RA   Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
RT   "The DCC gene: structural analysis and mutations in colorectal
RT   carcinomas.";
RL   Genomics 19:525-531(1994).
RN   [5]
RP   VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
RX   MEDLINE=94243823; PubMed=8187090;
RA   Miyake S., Nagai K., Yoshino K., Oto M., Endo M., Yuasa Y.;
RT   "Point mutations and allelic deletion of tumor suppressor gene DCC in
RT   human esophageal squamous cell carcinomas and their relation to
RT   metastasis.";
RL   Cancer Res. 54:3007-3010(1994).
CC   -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
CC       NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
CC   -!- DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
CC       DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
CC       EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
CC       POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
CC       METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   ------------------------------------------------------------------
```

```
DR   EMBL; X76132; CAA53735.1; -.
DR   EMBL; M32292; AAA35751.1; -.
DR   EMBL; M32286; AAA52174.1; -.
DR   EMBL; M32288; AAA52175.1; ALT_SEQ.
DR   EMBL; M32290; AAA52176.1; -.
DR   EMBL; M63696; AAA52177.1; -.
DR   EMBL; M63700; AAA52178.1; -.
DR   EMBL; M63702; AAA52179.1; -.
DR   EMBL; M63718; AAA52180.1; -.
DR   EMBL; M63698; AAA52181.1; -.
DR   PIR; A54100; A54100.
DR   PIR; A40098; A40098.
DR   PIR; A38442; A38442.
DR   HSSP; P56276; 1TLK.
DR   Genew; HGNC:2701; DCC.
DR   MIM; 120470;
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
KW   Repeat; Anti-oncogene; Disease mutation; Polymorphism;
KW   Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
FT   SIGNAL          1     25
FT   CHAIN          26   1447
FT   TRANSMEM     1098   1122   POTENTIAL.
FT   DOMAIN         26     54   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN        154    219   IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN        254    317   IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN        345    407   IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN        426    522   FIBRONECTIN TYPE-III 1.
FT   DOMAIN        525    618   FIBRONECTIN TYPE-III 2.
FT   DOMAIN        619    716   FIBRONECTIN TYPE-III 3.
FT   DOMAIN        722    816   FIBRONECTIN TYPE-III 4.
FT   DOMAIN        840    940   FIBRONECTIN TYPE-III 5.
FT   DOMAIN        941   1042   FIBRONECTIN TYPE-III 6.
FT   DISULFID       61    117   BY SIMILARITY.
FT   DISULFID      161    212   BY SIMILARITY.
FT   DISULFID      261    310   BY SIMILARITY.
FT   DISULFID      352    400   BY SIMILARITY.
FT   CARBOHYD       94     94   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      299    299   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      318    318   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      478    478   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      628    628   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      702    702   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      168    168   N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT       201    201   R -> G.
FT                              /FTId=VAR_003910.
FT   VARIANT      1375   1375   P -> H (IN A COLORECTAL CARCINOMA).
FT                              /FTId=VAR_003911.
FT   VARIANT       168    168   M -> T (IN OESOPHAGEAL SQUAMOUS CELL CARCINOMA).
FT                              /FTId=VAR_003909.
FT   CONFLICT      138    138   MISSING (IN REF. 3).
FT   CONFLICT      233    329   MISSING (IN REF. 3).
FT   CONFLICT      421    421   MISSING (IN REF. 3).
SQ   SEQUENCE   1447 AA;  158456 MW;  4A8612766ED0471F CRC64;
```

```
Query Match             6.2%;  Score 128;  DB 1;  Length 1447;
Best Local Similarity   24.8%;
Matches   60;  Conservative   33;  Mismatches   99;  Indels   50;  Gaps   10;
                                    Pred. No. 0.021;

Qy    10  SGSGNEV------------------IEGPQNATVLKGSQARFNCTVSQGW--KLIMWALSDM   51
             : ::|                   ||  | ::||    | |   | |   ||:|  | :
Db   220  SRTGNEAEVRILSDPGLHRQLYFLQRPSNVVAIEGKDAVLECCVS-GYPPPSFTWLRGEE  278

Qy    52  VVLSVRPEMPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSA  109
             |:|                ||  |    ||| |     ||| || | ||
Db   279  VI-------------QLRSKKYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNENISASA  323

Qy   110  YLTVQVMGELFIPSVNLLVVAAENEPCEVTCLPSHWTRLPDISW-ELGLLVSHSSSYFVPEP  168
          |  |      |  | :| ||  | |    |    | | | |||  || ||:      || |
Db   324  ELTVLVPPWFLNHPSNLYAYESMDIEFECTVS-GKPVPTVNWMKNGDVVIPSDYFQIVGG  382

Qy   169  SDLGSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLIVIRCPQDTGGINIPGVLSS  228
          |:| |                 | |          |    | |      |     :   |||
Db   383  SNLR----ILGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS------VLPS  430

Qy   229  LP  230
          ||
Db   431  AP  432
```

```
RESULT 6
KMLS_HUMAN
ID   KMLS_HUMAN    STANDARD;    PRT;   1914 AA.
AC   Q15746; Q14844; Q16794; O95796; O95797; O95798; O95799; Q9UBG5;
AC   Q9UIT9;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin light chain kinase, smooth muscle and non-muscle isozymes
DE   (MLCK) [Contains: Telokin (Kinase related protein)
DE   (KRP)].
DE   (EC 2.7.1.117)
GN   MYLK OR MLCK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Umbilical vein endothelial cells;
RX   MEDLINE=97304466; PubMed=9160829;
RA   Garcia J.G.N., Lazar V.L., Gilbert-Mcclain L.I., Gallagher P.J.,
RA   Verin A.D.;
```

```
RT   "Myosin light chain kinase in endothelium: molecular cloning and
RT   regulation.";
RL   Am. J. Respir. Cell Mol. Biol. 16:489-494(1997).
RN   [2]
RP   REVISIONS.
RA   Birukov K.G., Garcia J.G.N.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 2; 3A; 3B AND 4).
RC   TISSUE=Umbilical vein;
RX   MEDLINE=99216419; PubMed=10198165;
RA   Lazar V.L., Garcia J.G.N.;
RT   "A single human myosin light chain kinase gene (MLCK; MYLK).";
RL   Genomics 57:256-267(1999).
RN   [4]
RP   REVISIONS (ISOFORM 2).
RA   Birukov K.G., Garcia J.G.N.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 923-1914 FROM N.A.
RC   TISSUE=Hippocampus;
RX   MEDLINE=96121365; PubMed=8575746;
RA   Potier M.-C., Chelot E., Pekarsky Y., Gardiner K., Rossier J.,
RA   Turnell W.G.;
RT   "Analysis of the kinase-related protein gene found at human chromosome
RT   3q21 in a multi-gene cluster: organization, expression, alternative
RT   splicing and polymorphic marker.";
RL   J. Cell. Biochem. 75:481-491(1999).
RN   [6]
RP   SEQUENCE OF 1614-1914 FROM N.A.
RC   TISSUE=Lung, and Placenta;
RX   MEDLINE=20007838; PubMed=10536370;
RA   Watterson D.M., Schavocky J.P., Guo L., Weiss C., Chlenski A.,
RA   Shirinsky V.P., Van Eldik L.J., Haiech J.;
RT   "The human myosin light chain kinase (MLCK) from hippocampus:
RT   cloning, sequencing, expression, and localization to 3qcen-q21.";
RL   Genomics 29:562-570(1995).
RN   [7]
RP   SEQUENCE OF 1456-1914 FROM N.A.
RC   TISSUE=Placenta;
RA   Watterson M.D.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
CC       SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
CC       IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
CC       FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS. CENTRAL
CC       DETERMINANT IN THE DEVELOPMENT OF VASCULAR PERMEABILITY AND TISSUE
CC       EDEMA FORMATION. IN THE NERVOUS SYSTEM IT HAS BEEN SHOWN TO
CC       CONTROL THE GROWTH INITIATION OF ASTROCYTIC PROCESSES IN CULTURE
CC       AND TO PARTICIPATE IN TRANSMITTER RELEASE AT SYNAPSES FORMED
CC       BETWEEN CULTURED SYMPATHETIC GANGLION CELLS. CRITICAL PARTICIPANT
CC       IN SIGNALING SEQUENCES THAT RESULT IN FIBROBLAST APOPTOSIS.
CC   -!- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC       light-chain] phosphate.
CC   -!- SUBUNIT: TELOKIN BINDS CALMODULIN.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS: A NON-MUSCLE FORM (THE
CC       LONGEST FORM), A SMOOTH-MUSCLE FORM AND TELOKIN (A C-TERMINAL
CC       SECTION WITH NO CATALYTIC ACTIVITY); ARE PRODUCED BY ALTERNATIVE
CC       INITIATION.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS: 1 (SHOWN HERE), 2, 3A,
CC       3B, 4 AND DEL-1790; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES ARE
CC       EXPRESSED IN A WIDE VARIETY OF ADULT AND FETAL TISSUES AND IN
CC       CULTURED ENDOTHELIUM WITH QUALITATIVE EXPRESSION APPEARING TO BE
CC       NEITHER TISSUE- NOR DEVELOPMENT-SPECIFIC. NON-MUSCLE ISOFORM 2 IS
CC       THE DOMINANT SPLICE VARIANT EXPRESSED IN VARIOUS TISSUES. TELOKIN
CC       HAS BEEN FOUND IN A WIDE VARIETY OF ADULT AND FETAL TISSUES.
CC   -!- PTM: MLCK IS PROBABLY DOWN-REGULATED BY PHOSPHORYLATION.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
```

```
DR   EMBL; U48959; AAC18423.2; -.
DR   EMBL; AF069601; AAD15921.2; -.
DR   EMBL; AF069602; AAD15922.1; -.
DR   EMBL; AF069603; AAD15923.1; -.
DR   EMBL; AF069604; AAD15924.1; -.
DR   EMBL; X85337; CAA59685.1; -.
DR   EMBL; AF096771; AAD51380.1; JOINED.
DR   EMBL; AF096766; AAD51380.1; JOINED.
DR   EMBL; AF096767; AAD51380.1; JOINED.
DR   EMBL; AF096768; AAD51380.1; JOINED.
DR   EMBL; AF096769; AAD51380.1; JOINED.
DR   EMBL; AF096770; AAD51380.1; JOINED.
DR   EMBL; AF096774; AAD54018.1; -.
DR   EMBL; AF096771; AAD51381.1; -.
DR   EMBL; AF096769; AAD51381.1; JOINED.
DR   EMBL; AF096770; AAD51381.1; JOINED.
DR   HSSP; Q63450; 1A06.
DR   Genew; HGNC:7590; MYLK.
DR   MIM; 600922; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_like.
DR   InterPro; IPR003600; Ig_c2.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00047; ig; 8.
DR   PRINTS; PR00014; FNIYPEIII.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 8.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Serine/threonine-protein kinase; Calmodulin-binding;
KW   ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
KW   Alternative initiation; Alternative splicing.
FT   CHAIN            1   1914       MYOSIN LIGHT CHAIN KINASE, NON-MUSCLE
FT                                   ISOZYME.
FT   CHAIN          923   1914       MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE
FT                                   ISOZYME.
FT   INIT_MET      1761   1761       TELOKIN.
FT   INIT_MET       923    923       FOR MYOSIN LIGHT CHAIN KINASE, SMOOTH-
FT                                   MUSCLE ISOZYME.
FT   DOMAIN        1761   1761       FOR TELOKIN.
FT   DOMAIN        1343   1413       FIBRONECTIN TYPE-III.
FT   DOMAIN        1464   1719       PROTEIN KINASE.
FT   DOMAIN        1711   1774       CALMODULIN-BINDING.
FT   DOMAIN        1824   1891       IG-LIKE C2-TYPE DOMAIN.
FT   NP_BIND       1470   1478       ATP (BY SIMILARITY).
FT   BINDING       1493   1493       ATP (BY SIMILARITY).
FT   ACT_SITE      1585   1585       ATP (BY SIMILARITY).
FT   DOMAIN        1906   1914       POLY-GLU.
FT   REPEAT         868    895       5 X 28 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT         868    895       I-1.
FT   REPEAT         896    923       I-2.
FT   REPEAT         924    951       I-3.
FT   REPEAT         952    979       I-4.
FT   DOMAIN         980    998       I-5 (INCOMPLETE).
FT   DOMAIN         999   1063       6 X 12 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT         999   1003       II-1 (INCOMPLETE).
```

```
FT   REPEAT        1004   1015   II-2.
FT   REPEAT        1016   1027   II-3.
FT   REPEAT        1028   1039   II-4.
FT   REPEAT        1040   1051   II-5.
FT   REPEAT        1052   1063   II-6.
FT   VARSPLIC       437    506   VSGIPKPEVAWFLEGTPVRROEGSIEVYEDAGSHYICLLIKA
FT                                RTRDSGTYSCTASNAQGQVSCSWILQVER -> G (IN
FT                                ISOFORM 2 AND ISOFORM 3B).
FT   VARSPLIC      1433   1439   DEVEVSD -> MKWRCQT (IN ISOFORM 3A,
FT                                ISOFORM 3B AND ISOFORM 4).
FT   VARSPLIC      1473   1546   GKFGQVFRLVEKKTRKVWAGKFFKRAYSAKENIRQEISIM
FT                                NCLHHPKLVQCVDAFEEKANIVMVLEIVSGGEL -> L
FT                                (IN ISOFORM 4).
FT   CONFLICT      1655   1705   MISSING (IN ISOFORM 3A AND ISOFORM 3B).
FT   VARSPLIC      1790   1790   MISSING (IN ISOFORM DEL-1790).
FT   CONFLICT       933    933   V -> M (IN REF. 5).
FT   CONFLICT       963    963   S -> P (IN REF. 5).
FT   CONFLICT      1022   1022   P -> A (IN REF. 5).
FT   CONFLICT      1048   1050   KPM -> EAH (IN REF. 5).
FT   CONFLICT      1162   1162   P -> L (IN REF. 3; AAD15922).
FT   CONFLICT      1210   1210   L -> P (IN REF. 5).
FT   CONFLICT      1280   1280   A -> S (IN REF. 3; AAD15922/AAD15923).
FT   CONFLICT      1284   1284   M -> I (IN REF. 3; AAD15922/AAD15923/
FT                                AAD15924).
FT   CONFLICT      1300   1300   A -> G (IN REF. 5).
FT   CONFLICT      1316   1316   A -> G (IN REF. 5).
FT   CONFLICT      1326   1326   T -> S (IN REF. 5).
FT   CONFLICT      1478   1478   V -> C (IN REF. 5).
FT   CONFLICT      1511   1511   S -> T (IN REF. 5).
FT   CONFLICT      1563   1563   I -> T (IN REF. 5).
FT   CONFLICT      1609   1609   A -> P (IN REF. 5).
FT   CONFLICT      1639   1639   E -> D (IN REF. 3; AAD15922/AAD15923/
FT   CONFLICT      1640   1640   GY -> D (IN REF. 3; AAD15922/AAD15923/
```

```
Query Match  6.2%;  Score 128;  DB 1;  Length 1914;
Best Local Similarity  20.3%;  Pred. No. 0.03;
  Matches 86;  Conservative 58;  Mismatches 149;  Indels 130;  Gaps 19;

Qy    20 PQNATVLKGSQARFNCTVSGGW--KLIMWALSDMVVLSVRPMEPIITNDRFTSQRYDQGG 77
            ::|| ||: || |   ||||| |
Db    39 PRNLCIKEGATAKFEGRV-RGYPEPQVTWH------------RNGQPITSGGRFL---LDCGI 85

Qy    78 NFTSEMIIHNVEPSDSGNIRCSLQNSRLHSSAYLTVQV------------MGEL 119
            ::||| |
Db    86 RGTFSLVIHAVHEEDRGKYTCEATNG--SGARQVVTVELTVEGSFAKQLGQPVVSKTLGDR 143

Qy   120 FI-----------NLVVAENE-----PCEVTCLPSHWTRLPDISWELG- 154
            |:|
Db   144 FSASAVETRPSIWGEECPPKFATKLGRVVWKEGOMGRFSCKITIGRPQ----POVTWLKGN 198

Qy   155 LLVSHSSTYFVPEPSDLQSAVSTLALTPQSNGTLTCVATWKSLKARKSA----------TVN 206
            ||:
Db   199 VPLQPSARVSVSEKNGMQ-VLEIHGVNQDDVGVYTCLVVNGSGKASMSAELSIQGLDSAN 257

Qy   207 LTVIRCPQDTGGGT------NIPGVLSLSLPGFSLPTWGKVGLGLAGTMLLTPTCTLTIR 261
            :||
Db   258 RSFVRETKATNSDVRKEVTNVISKESKLDSL-------------------------- 288

Qy   262 CCCCRRRCCGCNCCCRCCFCCRRKRGRFRIQFOKKSEKEKTNKETETESGNEN---SGYNS 318
            :  |::         :::: ::  :
Db   289 EAAAKSKNCSSPQRGGSPPWAANSQPQPP-RESKLESCKDSPRTAPQTP 336

Qy   319 DEQKTTDTASLPPKSCESSDPEQRNSSCGPPHQRADQR-----PPRPASHPQASFNLASPE 374
            ::|::||   :||:| | :       :   ||:
Db   337 VLQKTSSSTILQAARVQ--PEPRAPGLVLSPSGEERRPAPPPRPATFPTRQPLGSQD 393

Qy   375 KVS 377
            |||
Db   394 VVS 396
```

RESULT 7

```
ID   DCC_MOUSE      STANDARD;      PRT;   1447 AA.
AC   P70211;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tumor suppressor protein DCC precursor.
GN   DCC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=96112625; PubMed=8570174;
RA   Cooper H.M., Armes P., Britto J., Gad J., Wilks A.F.;
RT   "Cloning of the mouse homologue of the deleted in colorectal cancer
RT   gene (mDcc) and its expression in the developing mouse embryo.";
RL   Oncogene 11:2243-2254(1995).
RN   [2]
RP   REVISIONS.
RC   STRAIN=BALB/c; TISSUE=Brain;
RA   Cooper H.M.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; A (shown here) and B; are
CC       produced by alternative initiation. A third isoform: C is
CC       produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
CC       DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
CC       BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC       ISOFORM C IS EXPRESSED ONLY IN THE EMBRYO.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS
CC       EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
CC       AND REMAIN AT THIS LEVEL IN THE ADULT.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
```

```
DR   EMBL; X85788; CAA59786.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:94869; Dcc.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
KW   Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW   Repeat; Anti-oncogene; Alternative initiation; Alternative splicing.
FT   SIGNAL          1     25   POTENTIAL.
FT   CHAIN          26   1447   TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM A.
FT   CHAIN          85   1447   TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM B.
FT   INIT_MET       85     85   FOR ISOFORM B.
FT   DOMAIN         26   1097   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     1098   1122   POTENTIAL.
FT   DOMAIN       1123   1447   CYTOPLASMIC (POTENTIAL).
```

```
FT   DOMAIN         54    124       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN        154    219       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN        254    317       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN        345    407       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN        426    522       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN        525    618       FIBRONECTIN TYPE-III 1.
FT   DOMAIN        619    716       FIBRONECTIN TYPE-III 2.
FT   DOMAIN        722    816       FIBRONECTIN TYPE-III 3.
FT   DOMAIN        840    940       FIBRONECTIN TYPE-III 4.
FT   DOMAIN        941   1042       FIBRONECTIN TYPE-III 5.
FT   DISULFID       61    117       BY SIMILARITY.
FT   DISULFID      161    212       BY SIMILARITY.
FT   DISULFID      261    310       BY SIMILARITY.
FT   DISULFID      352    400       BY SIMILARITY.
FT   CARBOHYD       60     60       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD       94     94       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD      299    299       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD      318    318       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD      478    478       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD      628    628       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD      702    702       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   VARSPLIC      819    838       MISSING (IN ISOFORM C).
FT   SEQUENCE     1447 AA;  158298 MW;  0D1F1097C22D5B9F CRC64;
```

```
Qy    17  IEGPQNATVLKGSQARFNCTVSQGW--KLIMWALSDMVVLSVRPMEPIITNDRFTSQRYD   74
              :|:|  ::|       |||||| |   |  |   :| |   ||: |  |  |:::
Db   243  LQRPSNVIAIEGKDAVLECCVS-GYPPPSFTWLRGEEVI------------QLRSKKYS   288

Qy    75  QGGNFTSEMIIHINVEPSDSGNIRC--SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAENE  132
              ||| || :|| | ||| | ||    |||| |||  |  |  |: |  | || |  ||
Db   289  LLGG--SNLIISNVTDDDSGTYTCVVTYKNENISASAELTVLVPPWFLNHPSNLYAYESM  346

Qy   133  PCEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCV  191
              :: |  |    ||   |   |:    ||  | ||| |:  |||   |||  |  ||
Db   347  DIEFECAVS-GKPVPTVNWMKNGDVVIPSDYFQIVGGSNLR----ILGVVKSDEGFYQCV  401

Qy   192  ATWKSLKARKSA  203
              ::  ||  ||
Db   402  AENEAGNAQSSA  413
```

```
Query Match    6.2%;  Score 127;  DB 1;  Length 1447;
Best Local Similarity   26.6%;
Matches  51;  Conservative  29;  Mismatches  86;  Indels  26;  Gaps  8;
         Pred. No. 0.026;
```

RESULT 8

```
PTK7_HUMAN
ID   PTK7_HUMAN     STANDARD;      PRT;   1070 AA.
AC   Q13308; Q13417;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase-like 7 precursor (Colon carcinoma kinase-4)
DE   (CCK-4).
GN   PTK7 OR CCK4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon carcinoma, and Placenta;
RX   MEDLINE=96074849; PubMed=7478540;
RA   Mossie K., Jallal B., Alves F., Sures I., Plowman G.D., Ullrich A.;
RT   "Colon carcinoma kinase-4 defines a new subclass of the receptor
RT   tyrosine kinase family.";
RL   Oncogene 11:2179-2184(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fibroblast;
RX   MEDLINE=97037064; PubMed=8882711;
RA   Park S.-K., Lee H.-S., Lee S.-T.;
RT   "Characterization of the human full-length PTK7 cDNA encoding a
RT   receptor protein tyrosine kinase-like molecule closely related to
RT   chick KLG.";
RL   J. Biochem. 119:235-239(1996).
RN   [3]
RP   SEQUENCE FROM N.A. AND REVISION TO 834.
RA   Lee S.-T., Park S.-K., Ji A.R., Jung J.W.;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY FUNCTION AS A CELL ADHESION MOLECULE. LACKS PROBABLY
CC       THE CATALYTIC ACTIVITY OF TYROSINE KINASE. MAY BE CONNECTED TO THE
CC       PATHOPHYSIOLOGY OF COLON CARCINOMAS AND/OR MAY REPRESENT A TUMOR
CC       PROGRESSION MARKER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, LIVER, PANCREAS,
CC       KIDNEY, PLACENTA AND MELANOCYTES. WEAKLY EXPRESSED IN THYROID
CC       GLAND, OVARY, BRAIN, HEART AND SKELETAL MUSCLE. ALSO EXPRESSED IN
CC       ERYTHROLEUKEMIA CELLS, BUT NOT EXPRESSED IN COLON.
CC   -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC       RECEPTOR SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U33635; AAA87555.1; -.
DR   EMBL; U40271; AAC50484.2; -.
DR   EMBL; AF447176; AAL39062.1; -.
DR   EMBL; AF447157; AAL39062.1; JOINED.
DR   EMBL; AF447158; AAL39062.1; JOINED.
DR   EMBL; AF447161; AAL39062.1; JOINED.
DR   EMBL; AF447164; AAL39062.1; JOINED.
DR   EMBL; AF447167; AAL39062.1; JOINED.
DR   EMBL; AF447170; AAL39062.1; JOINED.
DR   EMBL; AF447171; AAL39062.1; JOINED.
DR   EMBL; AF447173; AAL39062.1; JOINED.
DR   EMBL; AF447174; AAL39062.1; JOINED.
DR   EMBL; AF447175; AAL39062.1; JOINED.
DR   HSSP; P06631; 1AD5.
DR   Genew; HGNC:9618; PTK7.
DR   MIM; 601890;
DR   InterPro; IPR00719; Euk_pkinase.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig-like.
DR   InterPro; IPR002011; RTKinaseII.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 7.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD0000001; Euk_pkinase; 1.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 5.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00239; RECEPTOR_TYR_KIN_II; FALSE_NEG.
KW   Receptor; Transmembrane; Signal; Glycoprotein; Cell adhesion;
KW   Immunoglobulin domain; Repeat.
FT   SIGNAL          1     30       POTENTIAL.
FT   CHAIN          31   1070       TYROSINE-PROTEIN KINASE-LIKE 7.
FT   DOMAIN         31    704       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      705    725       POTENTIAL.
FT   DOMAIN        726   1070       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN         46    108       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN        143    208       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN        239    308       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN        336    398       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN        426    488       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN        517    577       IG-LIKE C2-TYPE DOMAIN 6.
```

```
FT   DOMAIN       606    671   IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN       796   1066   PROTEIN KINASE; INACTIVE.
FT   DISULFID      53    101   BY SIMILARITY.
FT   DISULFID     150    200   BY SIMILARITY.
FT   DISULFID     246    301   BY SIMILARITY.
FT   DISULFID     343    391   BY SIMILARITY.
FT   DISULFID     433    481   BY SIMILARITY.
FT   DISULFID     524    570   BY SIMILARITY.
FT   DISULFID     613    664   BY SIMILARITY.
FT   CARBOHYD     116    116   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     175    175   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     184    184   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     214    214   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     268    268   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     283    283   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     405    405   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     463    463   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     567    567   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     646    646   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD      92     92   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CONFLICT     147    147   P -> R (IN REF. 2 AND 3).
FT   CONFLICT     207    207   K -> T (IN REF. 2 AND 3).
FT   CONFLICT     495    496   S -> G (IN REF. 2 AND 3).
FT   CONFLICT     515    515   VL -> RV (IN REF. 2 AND 3).
FT   CONFLICT     881    881   G -> R (IN REF. 2 AND 3).
FT   CONFLICT     969    969   A -> P (IN REF. 2 AND 3).
FT   CONFLICT     992    992   S -> F (IN REF. 2 AND 3).
SQ   SEQUENCE   1070 AA;  118260 MW;  47CDF25B8E36498A5 CRC64;

Query Match             6.08%;  Score 124;  DB 1;  Length 1070;
  Best Local Similarity   25.3%;  Pred. No. 0.031;
  Matches   58;  Conservative   32;  Mismatches 103;  Indels   36;  Gaps   10;

Qy          16 VIEGPQNATVLKGSQARFNCTVS-OGWKLIMWALSDMVVLSVRPMEPIITNDRFTSQRYD 74
               ||  ||:||| | :|  | |:    |  |::| | ||:|| |:| ||:  |:  | :|
Db         227 VVLAPQDVVVARYEEAMFHCQFSAQPPPSLQWLFEDETPITNRSRPPHLRRATVPA---- 282

Qy          75 QGGNFTSEMIIHNVEPSDSGNIRCSLQNSR----LHGSAYLTVQVMGELFIPSVNLVVA 129
               :   :   ::    |  :::| |||           |||||||||  ||||||| |||
Db         283 -----NGSLLTQVRPRNAGIYRCIGQGQRGPPIILEATLHLAEIEDMPLFEPRVFTAGS 337

Qy         130 ENEPCEVTCLPSHWTRLPDISWE-LGL-LVSHSSYYFVPEPSDLQSAVSILALTPQSN-G 186
               | |  : | |   |  | :  |  :|| ::   :     |: |   | : |  :::  |
Db         338 EE---RVTCLPPKGLPEPSVWWEHAGVRLPTHGRV------QKGHELVLANIAESDAG 387

Qy         187 TLTCVATWKSLKARKSATVNLTV------IRCPQDTGGGINIPGVLSSL 229
               |:|| |  | |:|::   ::| ||       |||||  ||:|| |:  |
Db         388 VYTCHAA--NIAGQRRQDVNITVATVPSWLKKPQDSQLEEGKPGYLDCL 434
```

RESULT 9
OPCM_HUMAN

```
ID   OPCM_HUMAN     STANDARD;      PRT;   345 AA.
AC   Q14982;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Opioid binding protein/cell adhesion molecule precursor (OBCAM)
DE   (Opioid-binding cell adhesion molecule) (OPCML).
GN   OPCML OR OBCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Occipital cortex;
RX   MEDLINE=95237612; PubMed=7721093;
RA   Shark K.B., Lee N.M.;
RT   "Cloning, sequencing and localization to chromosome 11 of a cDNA
RT   encoding a human opioid-binding cell adhesion molecule (OBCAM).";
RL   Gene 155:213-217(1995).
CC   -!- FUNCTION: BINDS OPIOIDS IN THE PRESENCE OF ACIDIC LIPIDS; PROBABLY
CC       INVOLVED IN CELL CONTACT.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (BY
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ------------------------------------------------------------------------
DR   EMBL; L34774; AAA36387.1; -.
DR   Genew; HGNC:8143; OPCML.
DR   MIM; 600632; -.
DR   InterPro; IPR003006; Ig-MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IG2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL         1     27
FT   CHAIN         28    322   OPIOID BINDING PROTEIN/CELL ADHESION
FT                            MOLECULE.
FT   PROPEP       323    345   REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN        50    122   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN       150    209   IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN       237    303   IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID      57    115   POTENTIAL.
FT   DISULFID     157    202   POTENTIAL.
FT   DISULFID     244    296   POTENTIAL.
FT   CARBOHYD      44     44   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD      70     70   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     140    140   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     285    285   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     293    293   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     306    306   N-LINKED (GLCNAC...)  (POTENTIAL).
FT   LIPID        322    322   GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE    345 AA;  38007 MW;  E7AD17BEA1AA3FF4 CRC64;

Query Match             5.8%;  Score 120;  DB 1;  Length 345;
  Best Local Similarity   25.0%;  Pred. No. 0.016;
  Matches   71;  Conservative   37;  Mismatches 102;  Indels   74;  Gaps   18;

Qy           4 VIFLHGSG------SGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVL-- 54
               ::|| :|        |:|:|  ||:|| | :: |:| | |:   |::| | ||:|:|
Db          21 LILFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAGN 79

Qy          55 ---SVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSRL 105
                  |:   :|: :|    ||:     :  :  ::  :: || :  :|| |        |
Db          80 DKWSIDPRVIILVN---TPTQY----SIMIQNVDVDEGPYTCSVQTDNHFKTISRV 128

Qy         106 HGSAYLTVQVMGELFIPSVNLVVAAENEPCEVTCL----PSHWTRLPDISWELGLLVSHSS 161
               |   |::|| |  |:::|| ::  ::    : | |        ||| :    |: |: |
Db         129 H---LIVQVPPQIMNISSDITVNEGSSVILLCIAIGRPE------PTVTWR------HLS 173

Qy         162 YY-----FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-----IRC 212
                      :  :   :  :  :|: | :: ||| |:  :    |:: |:  :|      | |
Db         174 VKEGQGFVSEDEYLE----ISDIKRQSGEYECSAL-NDVAAPDVRKVKITVNYPPYISK 228

Qy         213 PQDTGGGINIPGVL-----SSLPSLGFSLPTWGK-----VGLGLAG 248
               |::|| :: |  :|      |:|  | |:|  |       ||:| |
Db         229 AKNTGVSVGQKGILSCEASAVPMAEFQ----WFKEETRLATGLDG 269
```

RESULT 10
OPCM_BOVIN

```
ID   OPCM_BOVIN     STANDARD;      PRT;   345 AA.
```

AC   P11834;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Opioid-binding protein/cell adhesion molecule precursor (OBCAM)
DE   (Opioid-binding cell adhesion molecule) (OPCML).
GN   OPCML OR OBCAM OR OCAM.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=89251576; PubMed=2721489;
RA   Schofield P.R., McFarland K.C., Hayflick J.S., Wilcox J.N.,
RA   Cho T.M., Roy S., Lee N.M., Loh H.H., Seeburg P.H.;
RT   "Molecular characterization of a new immunoglobulin superfamily
RT   protein with potential roles in opioid binding and cell contact.";
RL   EMBO J. 8:489-495(1989).
CC   -!- FUNCTION: BINDS OPIOIDS IN THE PRESENCE OF ACIDIC LIPIDS; PROBABLY
CC   INVOLVED IN CELL CONTACT.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC   similarity).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC   SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----
DR   EMBL; X12672; CAA31192.1; -.
DR   PIR; S03199; S03199.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL   1   27
FT   CHAIN   28   322   OPIOID BINDING PROTEIN/CELL ADHESION
FT   MOLECULE.
FT   PROPEP   323   345   REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN   50   122   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN   150   209   IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN   237   303   IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID   57   115   POTENTIAL.
FT   DISULFID   157   202   POTENTIAL.
FT   DISULFID   244   296   POTENTIAL.
FT   CARBOHYD   44   44   N-LINKED (GLCNAC...) (PROBABLE).
FT   CARBOHYD   70   70   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   140   140   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   285   285   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   293   293   N-LINKED (GLCNAC...) (PROBABLE).
FT   CARBOHYD   306   306   N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID   322   322   GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   345 AA; 37914 MW; D1ECC8D9E7D8C8B19 CRC64;

Qy   Query Match   5.8%;
     Best Local Similarity 25.1%;
     Matches   70; Conservative 36; Mismatches 103; Indels 70; Gaps 17;
                    Score 119; DB 1; Length 345;
                    Pred. No. 0.02;

```
Qy   4 VIFLHGSG----SGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVL---- 54
       :||:: |:  ::||  |||||  |||||  ||:||:|
Db  21 LIFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAGN 79
```

RESULT 11
OPCM_RAT
ID   OPCM_RAT   STANDARD;   PRT;   345 AA.
AC   P32736; Q01654; P32735; Q01653;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Opioid-binding protein/cell adhesion molecule precursor (OBCAM)
DE   (Opioid-binding cell adhesion molecule) (OPCML).
GN   OPCML OR OBCAM.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92347701; PubMed=1339369;
RA   Lippman D.A., Lee N.M., Loh H.H.;
RT   "Opioid-binding cell adhesion molecule (OBCAM)-related clones from a
RT   rat brain cDNA library.";
RL   Gene 117:249-254(1992).
RN   [2]
RP   SEQUENCE OF 195-214, AND GPI-ANCHOR.
RX   MEDLINE=95198094; PubMed=7891157;
RA   Struyk A.F., Canoll P.D., Wolfgang M.J., Rosen C.L., D'Eustachio P.,
RA   Salzer J.L.;
RT   "Cloning of neurotrimin defines a new subfamily of differentially
RT   expressed neural cell adhesion molecules.";
RL   J. Neurosci. 15:2141-2156(1995).
CC   -!- FUNCTION: BINDS OPIOIDS IN THE PRESENCE OF ACIDIC LIPIDS; PROBABLY
CC   INVOLVED IN CELL CONTACT.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC   MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC   SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----
DR   EMBL; M88710; AAA40859.1; -.
DR   EMBL; M88711; AAA40860.1; -.
DR   EMBL; M88709; AAA40858.1; -.
DR   PIR; JC1239; JC1239.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.

```
Qy  55 ---SVRMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSRL 105
         |:|: || ||  |  ::||  ||  |||:|||: | |:|  |:|
Db  80 DKWSIDPRVILLVN--TPFQY------SIMIQNVDVYDEGGPYTCSVQTDNHPKTSRV 128

Qy 106 HGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCL-----PSHWTRLPDISWELGLLVSHSS 161
        |:  ||| ||| ::  | |   :|:|  |:|       | |:| | ||||: |||  ||
Db 129 H----LIVQVPQIMNISSDVTVNEGSSVTLLCLAIGRPE---PTVTWR-----HLS 173

Qy 162 YY-----FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLIV------IRC 212
        :|:|     |: |  :||: |  :|  | | |||| |  |  |:|  |::|
Db 174 VKEGQGFVSEDEYLE----ISDIKRDQSGEYECSAL-NDVAAPDVRKVKITVNYNPPYISK 228

Qy 213 PQDIGGGINIPGVL-----SSLPSLGFSLPTWGKVYGLGLA 247
         ::|||   :   |     ::  |:  ::  | |:|| |:|
Db 229 AKNTGVSVGCQKGILSCEASAVPMAEFQ---WFKEDTRLA 264
```

```
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGC2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal; Alternative splicing.
FT   SIGNAL        1     27       BY SIMILARITY.
FT   CHAIN        28    322       OPIOID BINDING PROTEIN/CELL ADHESION
FT                                MOLECULE.
FT   PROPEP      323    345       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       50    122       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      150    209       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      237    303       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID     57    115       POTENTIAL.
FT   DISULFID    157    202       POTENTIAL.
FT   DISULFID    244    296       POTENTIAL.
FT   CARBOHYD     44     44       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     70     70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    306    306       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    322    322       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID         1             GPI-ANCHOR (POTENTIAL).
FT   VARSPLIC      1     27       MGVCGYLFLPWKCLVVVSLRLPFLVPT -> MYHPAVWIVF
FT                                SATTALLFIP (IN ISOFORM 2).
SQ   SEQUENCE    345 AA;  38067 MW;  A3I8IB0753F96658E CRC64;

Query Match             5.7%;
Best Local Similarity  24.7%;
Matches  71;  Conservative  38;  Mismatches 104;  Indels  74;  Gaps  18;

Qy    4  VIFLHGSG------SGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVL----  54
            :||:||||      :|||:|| ||||||||||||||||||||||||||||||||||
Db   21  LLFLVPTPGVPVRSGDATFPKAMDNVTVRQGESASATLRCTIDDRVTRVAW-LNRSTILYAGN  79

Qy   55  ---SVRMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ-------NSRL  105
            :|| :|| :||||
Db   80  DKWSIDRVIILVN--TPTQY---SIMIQNVDVDEGPYTCSVQTDNHPKTSRV  128

Qy  106  HGSAYLIVQVMGELFIPSVNLVVAENEPCEVTCL---PSHWTRLPDISWELGLLVSHSS  161
Db  129  H----LIVQVPPQIMNISSDIIVNEISSVTLLCIAIGRPE----PTVTWR----HLS  173

Qy  162  YY---FVPEPSDLQSAVSILALTPQSNGTLICVATWKSLKARKSATVNLLTV----IRC  212
Db  174  VKEGQGFVSEDEVLE----ISDIKRDQSGEYECSAL-NDVAAPDVRKVKITVVNYPPYISK  228

Qy  213  PQDTGGGINIPGVL---SSLPSLGFSLIPTWGK-----VGLGLAGTML  251
Db  229  AKNTGVSVGQKGILSCEASAVPMAEFQ------WFKEDTRLIATGLDGVRI  272
```

```
RESULT 12
NEO1_CHICK
ID   NEO1_CHICK              STANDARD;      PRT;   1443 AA.
AC   O90610;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neogenin (Fragment).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Embryonic brain;
RX   MEDLINE=95105243; PubMed=7806578;
RA   Vielmetter J., Roman J.M., Dreyer W.J.;
RT   "Neogenin, an avian cell surface protein expressed during terminal
RT   neuronal differentiation, is closely related to the human tumor
RT   suppressor molecule deleted in colorectal cancer.";
RL   J. Cell Biol. 127:2009-2020(1994).
CC   -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC       TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC       DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC       MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- DEVELOPMENTAL STAGE: IN RETINA, EXPRESSED ON GANGLION CELL FIBERS,
CC       AS SOON AS THEY BEGIN TO EXTEND THEIR AXONS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
KW   Transmembrane; Immunoglobulin domain; Glycoprotein; Repeat.
DR   EMBL; U07644; AAC59662.1; -.
DR   HSSP; P11276; 2MFN.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003598; Ig_MHC.
DR   InterPro; IPR003599; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 4.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGC2; 2.
FT   NON_TER       1      1
FT   DOMAIN       <1   1090       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1091   1111       POTENTIAL.
FT   DOMAIN     1112   1443       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       33    102       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      132    194       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      229    293       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      321    383       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      422    519       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      522    615       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      616    714       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      720    814       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      835    935       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      936   1037       FIBRONECTIN TYPE-III 6.
FT   DISULFID     40     95       BY SIMILARITY.
FT   DISULFID    139    187       BY SIMILARITY.
FT   DISULFID    236    286       BY SIMILARITY.
FT   DISULFID    328    376       BY SIMILARITY.
FT   CARBOHYD     39     39       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    176    176       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    456    456       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    475    475       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    625    625       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    700    700       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    894    894       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   1443 AA;  158050 MW;  558C679557 9C0E26 CRC64;

Query Match             5.7%;
Best Local Similarity  26.28%;
Matches  49;  Conservative  28;  Mismatches  73;  Indels  37;  Gaps  11;

Qy   17  IEGPQNATVLKGSQARFNCTVSQGW-KLIMWALSDMVVLSVRPMEPIITNDRFTSQRY-  73
            ||:    ||:|||:||||||||||:|: |:||
Db  218  VRQPSSLRKVTGQNAVFPC-VAGGFPTPVVRWTKNG------EELIFED--SERFA  264

Qy   74  -DQGGNFTSEMIIHNVEPSDSGNIRC---SLQNSRLHGSAYLIVQVMGELFIPSVNLVVAE  130
            ||: ||| :||
Db  265  LIRAGGS-----LILISDVTEEDVGIYTCIADNENETIEAQAELAVQVPPEFLKRPANIYAHE  320
```

```
Qy  131 NE----PCEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSN 185
            ||||  |:||   |:|  | |        |:| |:| |:|  |
Db  321 SMDIVFECEVTGKPT-----PTVKWVKNGDVVIPSDYFKIVKEHNLQ----VLGIVKSDE 371

Qy  186 GTLTCVA 192
            | : |
Db  372 GFYQCIA 378
```

RESULT 13

```
ID   PGBM_HUMAN     STANDARD;      PRT;   4393 AA.
AC   P98160; Q16287;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin, and Colon;
RX   MEDLINE=92235084; PubMed=1569102;
RA   Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. Biol. Chem. 267:8544-8557(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92112994; PubMed=1730768;
RA   Kallunki P., Tryggvason K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. Cell Biol. 116:559-571(1992).
RN   [3]
RP   SEQUENCE OF 1018-1472 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91365376; PubMed=1679749;
RA   Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA   Yi H.F., Iozzo R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   Genomics 10:673-680(1991).
RN   [4]
RP   SEQUENCE OF 892-1398 FROM N.A.
RC   TISSUE=Fibrosarcoma;
RX   MEDLINE=92120660; PubMed=1685141;
RA   Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA   Tryggvason K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1-->p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   Genomics 11:389-396(1991).
RN   [5]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=94052171; PubMed=8234307;
RA   Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC   -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC   -!- SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
```

```
DR   EMBL; X62515; CAA44373.1; -.
DR   EMBL; M85289; AAA52700.1; -.
DR   EMBL; M64283; AAA52699.1; -.
DR   EMBL; S76436; AAB21121.2; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   HSSP; P00740; 1EDM.
DR   Siena-2DPAGE; P98160; -.
DR   Genew; HGNC:5273; HSPG2.
DR   MIM; 142461; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR000034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; ig; 22.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 3.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00181; EGF_Lam; 6.
DR   SMART; SM00180; EGF-like; 8.
DR   SMART; SM00408; IGc2; 22.
DR   SMART; SM00409; IG; 22.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00281; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00210; Laminin_EGF.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01248; EGF_LAM; 11.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; EGF-like domain.
FT   SIGNAL        1     21       POTENTIAL.
FT   CHAIN        22   4393       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                                SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN       80    194       SEA.
```

RESULT 14

NEO1_HUMAN

```
Qy   73  YDQGGNFTSEMIIHNVEPSDSGNIRC----------------SLQNSRLHGSAVLTVQVWGELFI-  121
             |:|| | |                        :|| :|.| |||
Db 3080  ------------------IVGTRPSNHGTYRCVASNAYGVAQSVVNLSVGGPPTVSVLPEGPVWVK  3127

Qy  122  --PSVNL-VVAENEPCEVTCLPSHWTRLPDISWEL-----GLLVSHSSYYFVPEPSDLQS  173
                |                                    |||
Db 3128  VGKAVTLECVSAGEPRS-----SARWTRISSTPAKLEQRIYGLMDSH------------T  3170

Qy  174  AVSILALTPQSNGTLTCVAATWKSLKARKSATVNLIIVIRCPQDTG  217
             :|                            |:|.| |||
Db 3171  VLQISSAKPSDAGTYVCLAQNALGTAQKQVEVIV-----DTG  3207
```

ID   NEO1_HUMAN   STANDARD;   PRT;   1461 AA.
AC   Q92859; O00340;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neogenin precursor.
GN   NEO1 OR NGN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97312699; PubMed=9169140;
RA   Vielmetter J., Chen X.-N., Miskevich F., Lane R.P., Yamakawa K.,
RA   Korenberg J.R., Dreyer W.J.;
RT   "Molecular characterization of human neogenin, a DCC-related protein,
RT   and the mapping of its gene (NEO1) to chromosomal position 15q22.3-
RT   q23.";
RL   Genomics 41:414-421(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97236653; PubMed=9121761;
RA   Meyerhardt J.A., Look A.T., Bigner S.H., Fearon E.R.;
RT   "Identification and characterization of neogenin, a DCC-related
RT   gene.";
RL   Oncogene 14:1129-1136(1997).
CC   -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC   TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC   DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC   MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC   ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AND ALSO IN CANCER CELL
CC   LINES.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC   SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.

DR   EMBL; U61262; AAB17263.1; -.
DR   EMBL; U72391; AAC51287.1; -.
DR   HSSP; P02751; 1TTF.
DR   Genew; HGNC; 7754; NEO1.
DR   MIM; 601907; -.
DR   InterPro; IPR003961; FN_III.

FT   DOMAIN   197   236   LDL-RECEPTOR CLASS A 1.
FT   DOMAIN   283   321   LDL-RECEPTOR CLASS A 2.
FT   DOMAIN   323   361   LDL-RECEPTOR CLASS A 3.
FT   DOMAIN   366   405   LDL-RECEPTOR CLASS A 4.
FT   DOMAIN   405   506   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN   523   532   LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN   533   732   LAMININ DOMAIN IV.1 (DOMAIN III A).
FT   DOMAIN   733   765   LAMININ EGF-LIKE 2.
FT   DOMAIN   766   815   LAMININ EGF-LIKE 3.
FT   DOMAIN   816   873   LAMININ EGF-LIKE 4.
FT   DOMAIN   881   925   LAMININ EGF-LIKE 5 (INCOMPLETE).
FT   DOMAIN   926   935   LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN   936   1127   LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN   1128   1160   LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN   1161   1210   LAMININ EGF-LIKE 6.
FT   DOMAIN   1211   1267   LAMININ EGF-LIKE 7.
FT   DOMAIN   1277   1326   LAMININ EGF-LIKE 8.
FT   DOMAIN   1327   1336   LAMININ DOMAIN IV 3 (N-TERMINAL).
FT   DOMAIN   1337   1531   LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN   1532   1564   LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN   1565   1614   LAMININ EGF-LIKE 10.
FT   DOMAIN   1615   1672   LAMININ EGF-LIKE 11.
FT   DOMAIN   1679   1773   IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN   1774   1867   IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN   1868   1957   IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN   1958   2053   IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN   2054   2153   IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN   2154   2246   IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN   2247   2342   IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN   2343   2438   IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN   2439   2535   IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN   2536   2631   IG-LIKE C2-TYPE DOMAIN 11.
FT   DOMAIN   2632   2728   IG-LIKE C2-TYPE DOMAIN 12.
FT   DOMAIN   2729   2828   IG-LIKE C2-TYPE DOMAIN 13.
FT   DOMAIN   2829   2926   IG-LIKE C2-TYPE DOMAIN 14.
FT   DOMAIN   2927   3023   IG-LIKE C2-TYPE DOMAIN 15.
FT   DOMAIN   3024   3114   IG-LIKE C2-TYPE DOMAIN 16.
FT   DOMAIN   3115   3213   IG-LIKE C2-TYPE DOMAIN 17.
FT   DOMAIN   3214   3300   IG-LIKE C2-TYPE DOMAIN 18.
FT   DOMAIN   3301   3401   IG-LIKE C2-TYPE DOMAIN 19.
FT   DOMAIN   3402   3490   IG-LIKE C2-TYPE DOMAIN 20.
FT   DOMAIN   3491   3576   IG-LIKE C2-TYPE DOMAIN 21.
FT   DOMAIN   3577   3664   IG-LIKE C2-TYPE DOMAIN 22.
FT   DOMAIN   3665   3845   LAMININ G-LIKE 1.
FT   DOMAIN   3846   3883   EGF-LIKE 1.
FT   DOMAIN   3886   3924   EGF-LIKE 2.
FT   DOMAIN   3930   4105   LAMININ G-LIKE 2.
FT   DOMAIN   4106   4143   EGF-LIKE 3.
FT   DOMAIN   4145   4178   EGF-LIKE 4.
FT   DOMAIN   4203   4391   LAMININ G-LIKE 3.
FT   DOMAIN   65   67   HEPARAN SULFATE (POTENTIAL).
FT   SITE   71   73   HEPARAN SULFATE (POTENTIAL).
FT   SITE   76   78   HEPARAN SULFATE (POTENTIAL).
FT   SITE   4151   4153   MEDIATES MOTOR NEURON ATTACHMENT
FT   (POTENTIAL).
FT   SITE   4301   4303   MEDIATES MOTOR NEURON ATTACHMENT
FT   (POTENTIAL).
FT   DISULFID   199   212   BY SIMILARITY.
FT   DISULFID   206   225   BY SIMILARITY.
FT   DISULFID   219   234   BY SIMILARITY.
FT   DISULFID   285   297   BY SIMILARITY.
FT   DISULFID   292   310   BY SIMILARITY.
FT   DISULFID   304   319   BY SIMILARITY.
FT   DISULFID   325   337   BY SIMILARITY.
FT   DISULFID   332   350   BY SIMILARITY.

Query Match 24.6%; Score 116; DB 1; Length 4393;
Best Local Similarity 5.6%;
Matches 55; Conservative 25; Mismatches 72; Indels 72; Gaps 10;

```
Qy   20  PQNATVLKGSQARFNCTVSQG----------WKLIMWALSDMVVLSVRPMEPIITNDRFTSQR  72
             :|| | ||  :  ||  :| |                                    | |
Db 3030  PPSSTVQQGQDASFKCLIHDGAAPISLEWKTRNQELEDNVHIS--PNGSITT----------  3079
```

```
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 3.
KW   Transmembrane; Immunoglobulin domain; Glycoprotein; Signal; Repeat;
KW   Alternative splicing.
FT   SIGNAL        1     33
FT   CHAIN        34   1461       NEOGENIN.
FT   DOMAIN       34   1105       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1106   1126       POTENTIAL.
FT   DOMAIN     1127   1461       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       67    136       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      166    228       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      263    327       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      355    417       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      436    533       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      536    629       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      630    729       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      735    829       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      850    950       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      951   1052       FIBRONECTIN TYPE-III 6.
FT   DOMAIN     1118   1121       POLY-VAL.
FT   DISULFID     74    129       BY SIMILARITY.
FT   DISULFID    173    221       BY SIMILARITY.
FT   DISULFID    270    320       BY SIMILARITY.
FT   DISULFID    362    410       BY SIMILARITY.
FT   CARBOHYD     73     73       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    210    210       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    326    326       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    639    639       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    715    715       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    909    909       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   VARSPLIC   1248   1300       MISSING (IN ISOFORM 2).
FT   CONFLICT    168    168       G -> N (IN REF. 2).
SQ   SEQUENCE   1461 AA;  159958 MW;  7AAE897E69635A21 CRC64;

Query Match         5.6%;
Best Local Similarity   24.4%;
Matches   68;  Conservative   31;  Mismatches  107;  Indels   73;  Gaps   13;

Qy    12 SGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITN------ 65
         :|| |  |                  |:||       |    |     ::|
Db   203 SGMLVI---SNAT--EGDGGLYRCVVESGGPP--KYSDEVELKVLPDPEVISDLVFLKQ 254

Qy    66 -------------------DRETSQRYD--QGGNFTSEMIHHNV 88
                            |:|| |   ||||:||||||
Db   255 PSPLVRVIGQDVVLPCVASGLPTPTIKWMKNEEALDTESSERIVLLAGGS---LEISDV 310

Qy    89 EPSDSGNIRCSLQNSR---LHGSAYLTVQGVMGELFIPSVNLVVAENE--- 142
         |:|    | |  :|       |||| :  ||||  |||||||| ||
Db   311 TEDDAGTYFCIADNGNETIEAQAELTVQQAQEFLKQPTNIYAHESMDIVFECEVTGKPT- 369

Qy   143 WTRLPDISW-ELGLLVSHSSYYFVPEFSDLQSAVSILALTFQQSNGTLITCVATWKSLKARK 201
         |  | |  |         |  :  |||||                |||  |: :
Db   370 ----PTVKWVKNGDMVIPSDYEKIVKEHNLQ----VLGLVKSDEGFYQCIAENDVGNAQA 421

Qy   202 SATVNLIVTVIRCPQDTGGGINIP-GVLSSLPSLGFSLPTW 239
                                  |:|::::|||:|||::|
Db   422 GAQL-IILEHAPTTGPLPSAPRDVVASLVSTRFIKLTW 459
```

```
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Advanced glycosylation end product-specific receptor precursor
DE   (Receptor for advanced glycosylation end products).
GN   AGER OR RAGE.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Lung;
RX   MEDLINE=92340547; PubMed=1378843;
RA   Neeper M., Schmidt A.M., Brett J., Yan S.D., Wang F., Pan Y.C.,
RA   Elliston K., Stern D., Shaw A.;
RT   "Cloning and expression of a cell surface receptor for advanced
RT   glycosylation end products of proteins.";
RL   J. Biol. Chem. 267:14998-15004(1992).
CC   -!- FUNCTION: MEDIATES INTERACTIONS OF ADVANCED GLYCOSYLATION END
CC       PRODUCTS (AGE). THESE ARE NONENZYMATICALLY GLYCOSYLATED PROTEINS
CC       WHICH ACCUMULATE IN VASCULAR TISSUE IN AGING AND AT AN ACCELERATED
CC       RATE IN DIABETES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: ENDOTHELIAL CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M91212; AAA03575.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Immunoglobulin domain; Glycoprotein; Transmembrane; Repeat; Signal.
FT   SIGNAL        1     22
FT   CHAIN        23    416       ADVANCED GLYCOSYLATION END PRODUCT-
FT                                SPECIFIC RECEPTOR.
FT   DOMAIN       23    352       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    353    373       POTENTIAL.
FT   DOMAIN      374    416       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       31    105       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      136    214       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      262    318       IG-LIKE C2-TYPE DOMAIN 2.
FT   DISULFID     38     98       POTENTIAL.
FT   DISULFID    143    207       POTENTIAL.
FT   DISULFID    269    311       POTENTIAL.
FT   CARBOHYD     25     25       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     80     80       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   DOMAIN      391    396       POLY-GLU.
SQ   SEQUENCE    416 AA;  44182 MW;  B703815573E767AE CRC64;

RESULT 15
RAGE_BOVIN
ID   RAGE_BOVIN     STANDARD;      PRT;   416 AA.
AC   Q28173;

Query Match         5.8%;
Best Local Similarity   23.1%;
Matches   75;  Conservative   35;  Mismatches   87;  Indels  127;  Gaps   18;

Qy    70 SQRYDQGGNFT--SEMIHHNVEPSDSGNIR----CS---LQNSRLHGSAYLTVQVMGE- 118
         :::| |||| |   || :|  :|||| ||
Db   176 TKRHPKIGLFTLHSELM--VTPARGGALHPFSCGSFPTPGLPRRALHTAPIQLRVWSEH 232

Qy   119 ---------LFIPSVNLVVAENEP------------------CEVTCLPS---HWTR---LP 147
```

```
Db  233  RGGEGPNVDAVPLKEVQLVV---EPEGGAVAPGGTVTLTCEAPAQPPQIHWIKDGRPLP  289
             :  |||     ||
Qy  148  DISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNL  207
             |:  |:|||||    :|:  |:
Db  290  -------------LPPGPMLLLPEVG-----PEDQGTYSCVATHPSHGPQESRAVSV  328
Qy  208  TVIRCPQD--TGGGINIPGVLSSLPSIGFSSLPTWGKVGLGLAGTMLLTPTCTLITRCCCC  265
             |:|   ::|         |:|          ||
Db  329  TLIETGEEGTTPAGSVEGPG-----LETLALTLGILG--GLG-----TVALLIGVIW  373
Qy  266  RRRCCGCGNCCCRCCFCCRRKRGFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKTTD  325
             |:|                        |:|
Db  374  HRR-------------------QERKVPENQEEEEERAELNQPE------  404
Qy  326  TASLPPKSCESSDPEQRNSSCGPP  349
             :|| |||:|
Db  405  -----------EPEAAESSTGGP  416
```

OM protein - protein search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.4_p5_4578

Run on: April 28, 2003, 18:09:54 ; Search time 18.8683 Seconds

US-09-729-264-2
(without alignments)
1946.300 Million cell updates/sec

Title:
Perfect score: 2059
Sequence: 1 MGLVIFLHGSGSGNEVIEGP............HPQASFNLASPEKVSNTTVV 382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	147.5	7.2	404	1	I61596	advanced glycosyla
2	133	6.5	1427	2	I51669	tumor suppressor -
3	129.5	6.3	164	2	T2427	hypothetical prote
4	129.5	6.3	344	2	I56551	neurotrimin - rat
5	128	6.2	1447	2	A54100	tumor suppressor p
6	127.5	6.2	871	2	I48696	protein-tyrosine k
7	127.5	6.2	881	1	I48697	protein-tyrosine k
8	125.5	6.1	6642	2	T29757	protein UNC-89 - C
9	125	6.0	1070	2	JC4593	protein-tyrosine k
10	124	5.9	188	2	T15651	hypothetical prote
11	122.5	5.9	188	2	JC4025	opioid-binding cel
12	120	5.8	345	2	S03199	opioid-binding pro
13	119	5.8	345	2	JC1239	opioid-binding pro
14	118	5.7	345	2	I50600	neogenin - chicken
15	117.5	5.7	338	2	JC1238	opioid-binding pro
16	117.5	5.7	4391	2	A38096	perlecan precursor
17	116	5.6	152	2	T18975	hypothetical prote
18	114.5	5.6	416	1	A42879	advanced glycosyla
19	113.5	5.5	764	2	A49448	irregular chiasm C
20	112	5.5	693	2	S49228	sodium-dependent p
21	111	5.4	620	2	JH0593	Schwann cell myeli
22	110.5	5.4	1177	2	T16594	hypothetical prote
23	110.5	5.4	2491	1	A28372	insulin-like growt
24	109.5	5.3	391	2	T09058	butyrophilin homol
25	109	5.3	487	2	S65133	butyrophilin - mou
26	108.5	5.3	882	2	I38912	receptor tyrosine
27	107	5.2	3707	2	S18252	heparan sulfate pr
28	106.5	5.2	662	2	S18251	hypothetical prote
29	106.5	5.2	802	2	T13149	mitogen-and stress
30	106.5	5.2	946	1	A47299	ror-related recept
31	106	5.1	1092	1	JN0635	neural cell adhesi
32	104.5	5.1	862	2	I49583	differentiation an
33	104	5.1	2222	2	T13924	sdk protein - frui
34	104	5.1	5825	2	T12117	polyprotein - fava
35	103.5	5.0	890	1	A53743	protein-tyrosine k
36	102.5	5.0	333	2	A31923	amalgam protein pr
37	102.5	5.0	423	2	T29549	hypothetical prote
38	102.5	5.0	789	2	T28714	hypothetical prote
39	102.5	5.0	1273	2	T42405	sax-3 protein - Ca
40	102.5	5.0	1355	2	T28715	hypothetical prote
41	102.5	5.0	2499	2	A30788	mannose 6-phosphat
42	102	5.0	365	1	JC7780	coxsackie- and ade
43	102	5.0	721	2	T41530	hypothetical prote
44	102	5.0	868	2	A46512	CD22 homolog/B lym
45	102	5.0*	880	1	JC4166	protein-tyrosine k

ALIGNMENTS

RESULT 1
I61596
advanced glycosylation end-products receptor precursor - human
advanced glycosylation end-products receptor precursor binding protein, 35K; glycoprot
N;Alternate names: Alternate names: advanced glycosylation end product-binding protein,
C;Species: Homo sapiens (man)
C;Date: 24-May-1996; #sequence_revision 07-Feb-1997 #text_change 16-Jul-1999
C;Accession: I61596; B42879; S27968
R;Sugaya, K.; Fukagawa, T.; Matsumoto, K.; Mita, K.; Takahashi, E.; Ando, A.; Inoko,
Genomics 23, 408-419, 1994
A;Title: Three genes in the human MHC class III region near the junction with the cla
nterpart of mouse mammary tumor gene int-3.
A;Cross-references: GB:D28769; NID:g561657; PIDN:BAA05958.1; PID:g561659
R;Neeper, M.; Schmidt, A.M.; Brett, J.; Yan, S.D.; Wang, F.; Pan, Y.C.; Elliston, K.;
J. Biol. Chem. 267, 14998-15004, 1992
A;Title: Cloning and expression of a cell surface receptor for advanced glycosylation
A;Reference number: A42879; MUID:92340547; PMID:1378843
A;Accession: A42879
A;Molecule type: DNA
A;Residues: 1-404 <RES>
A;Cross-references: GB:D28769; NID:g561657; PIDN:BAA05958.1; PID:g561659
A;Reference number: I61596
A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/
A;Molecule type: mRNA
A;Residues: 'G',2-99,'R',101-404 <NEE>
A;Experimental source: lung
A;Cross-references: EMBL:M91211; NID:g190845; PIDN:AAA03574.1; PID:g190846
A;Note: sequence extracted from NCBI backbone (NCBIP:109438)
C;Comment: Advanced glycosylation end products are heterogeneous nonenzymatically gly
cellular function, thus contributing to tissue lesions in diabetes.
C;Comment: This receptor appears also to mediate the effects of amyloid beta peptide
ates in the neurotoxic pathway that produces dementia in Alzheimer's disease.
C;Genetics:
A;Gene: GDB:AGER
A;Cross-references: GDB:306354; OMIM:600214
A;Map position: 6p21.3-6p21.3
A;Introns: 18/1; 53/3; 119/1; 140/3; 170/1; 231/1; 274/3; 322/1; 331/1; 373/2
A;Function:
C;Description: neuronal receptor for amphoterin, a DNA-binding protein involved in ne
C;Superfamily: advanced glycosylation end products receptor; immunoglobulin homology
C;Keywords: Alzheimer's disease; glycoprotein; receptor; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-404/Product: advanced glycosylation end products receptor #status predicted <MAT
F;23-344/Domain: extracellular #status predicted <EXT>
F;31-101/Domain: immunoglobulin homology <IM1>
F;137-210/Domain: immunoglobulin homology <IM2>
F;252-303/Domain: immunoglobulin homology <IM3>
F;345-362/Domain: transmembrane #status predicted <TMM>
F;363-404/Domain: intracellular #status predicted <INT>
F;25,81/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;38-99,144-208,259-301/Disulfide bonds: #status predicted

Query Match   7.2%;   Score 147.5;   DB 1;   Length 404;

```
RESULT 2
I51669
tumor suppressor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996
C;Accession: I51669
R;Pierceall, W.E.; Reale, M.A.; Candia, A.F.; Wright, C.V.; Cho, K.R.; Fearon, E.R.
A;Title: Expression of a homologue of the deleted in colorectal cancer (DCC) gene in the
Dev. Biol. 166, 654-665, 1994
A;Reference number: I51668; MUID:95113183; PMID:7813784
A;Accession: I51669
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1427 <PIE>
A;Cross-references: EMBL:U10986; NID:g606873; PIDN:AAA70168.1; PID:g606874
A;Genetics:
A;Gene: XDCca

Query Match       6.5%;  Score 133;  DB 2;  Length 1427;
Best Local Similarity   25.7%;  Pred. No. 0.088;
Matches   59;  Conservative   33;  Mismatches   90;  Indels   48;  Gaps   13;

Qy   17 IEGPQNATVLKGSQARFNCTVSQGW--KLIMWALSDMVVLSVRPMEPIITNDRFTS-  70
              ::| | ||||:|   :||    |:    :|| ||||:|:|: |::
Db  243 LQRPSNVVAIEGQDAVLECAVS-GYPTPTIVWMQGD------------RFTSQRYD  74

Qy   75 QGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAAENE  132
           ||| |||||:|||||||||||||   :|| |||||||||||:|||| |:|:| :|||
Db  289 VLGG--SNLLISNVTDDDAGAYTCVATVKNENTSFSADLTVMVPPQFLNHPANLYAYESM  346

Qy  133 PCEVTCL--PSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGT  187
         |||||||  |||:|||||||| |||||||||||||||||||:||||||||||||||||
Db  347 DIEFCAVSGKPS------PTVKWTKNGEVVIPSDYFQIVDGSNLR---ILGLVKSDEGY  397

Qy  188 LTCVA-----TWKSLKARKSATVNLIVI-RCPQDTGGGINIPGVLSS  228
         ||||      |||||||||||||||:|| :|:||||||||||:||||
Db  398 YQCIAENEAGNIQTYAQLIIPDPAVPSSSILPSAPRDV-----VPVLVSS  442
```

```
RESULT 3
T24272
hypothetical protein T01B7.8 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24272
R;Sims, M.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z19867
A;Accession: T24272
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-164 <WIL>
A;Cross-references: EMBL:Z66499; PIDN:CAA91301.1; GSPDB:GN000020; CESP:T01B7.8
A;Experimental source: clone T01B7
A;Genetics:
A;Gene: CESP:T01B7.8
A;Map position: 2
A;Introns: 20/3; 90/2

Query Match       6.3%;  Score 129.5;  DB 2;  Length 164;
Best Local Similarity   32.1%;  Pred. No. 0.015;
Matches   53;  Conservative   11;  Mismatches   53;  Indels   27;  Gaps   6;

Qy  153 LGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRC  212
         |  || ||: ||:    |:|::|||||||||:|||||:|| | ||||||||||| ||:|
Db    6 LAILLAIGTFIAV--SQVQSAV-----LPVSSTELATVGTDVSTASTAIDTLGNSSSRV  57

Qy  213 PQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLLAGTMLLLTPTCTLTLTIRCCCCRRRCCGC  272
         |:||||||||||||||||||||||||| |||||:|| ||:|||||||||||||||||||| |||
Db   58 KRQGGCGCGGGC------------------GCCGCGGGGGG--CGCCCCRPCCCCCRRCCTC  101

Qy  273 -NCCC-RCCFCCR  283
          ||| |||| ||
Db  102 CRTCCTRCCTCCR  115
```

```
RESULT 4
I56551
neurotrimin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 19-May-2000
C;Accession: I56551
R;Struyk, A.F.; Canoll, P.D.; Wolfgang, M.J.; Rosen, C.L.; D'Eustachio, P.; Salzer, J
J. Neurosci. 15, 2141-2156, 1995
A;Title: Cloning of neurotrimin defines a new subfamily of differentially expressed n
A;Reference number: I56551; MUID:95198094; PMID:7891157
A;Accession: I56551
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-344 <RES>
A;Cross-references: EMBL:U16845; NID:g755184; PIDN:AAA67445.1; PID:g755185
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-ter

Query Match       6.3%;  Score 129.5;  DB 2;  Length 344;
Best Local Similarity   24.2%;  Pred. No. 0.034;
Matches   64;  Conservative   43;  Mismatches   104;  Indels   53;  Gaps   14;

Qy    4 VIFLHGSG-----SGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLS---  55
         :| |||  | :|: |:||: ||||||||||||||||||||||||||||||||||||:
Db   21 LLFLIVPTGVPVRSGDATFPKAMDNVTVRQESAPLRCTIDNRVTRVAWLNRSTPILYAGND  80

Qy   56 --VRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSRLH  106
           |||||||||||||||:|    || ||||||||||||||||||||:|      ||:|:
Db   81 KWCLDPRVLLSN--TQYQY------SIEIQNVDYDEGPYTCSVQTDNHPKYSRVH  129

Qy  107 GSAYLTVQVMGELFIPSVNLVVAAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVP  166
         ||||||||:||||| ||||||||||||||  |||||:|||||:|| :|||||||||||||
Db  130 ----LIVQVSPKIVEISDSDISINEGNNILSTCIATGRPE-PTVTWR---HISPKAVGFVS  181

Qy  167 EPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP-----QDTGGGIN  221
         ||:|||||||||||||||||||||||||||||||||||||||||||      |:||||||
Db  182 EDEYLE----IQGITREQSGEYECSAS-NDVAAPVVRRVANVTNVNPPYISEAKGTGVPVG  236

Qy  222 IPGVL-----SSLPSLGFSLPTWGK  241
```

```
RESULT 1

Best Local Similarity   22.6%;  Pred. No. 0.0019;
Matches   85;  Conservative   40;  Mismatches   114;  Indels   137;  Gaps   18;

Qy   15 EVIEGPQNATVLKGSQARFNCTVSQG---WKLIMWALSDMVVLSVRPMEPIITNDRFTS-  70
         ::| | ||||:|:   :||   |     ::|| ||||:|:|: ||::      |: :|
Db  125 EIVDSASELTA--GVPNKVGTCVSEGSYPAGTLSWHLDG------------KPLVPNEKGVSV  173

Qy   71 ---QRYDQGGNFT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGSAYLTVQVMGELF  120
                 :|||||    ||:|:|||||||||||    ||||:||||||||||::|| |||
Db  174 KEQTRHPEFGLFTLQSELM---VTPARGGDRPTFSCSFSPGLPHRALRTAPIQPRVW  230

Qy  121 IP---SVNLVVAENEP------------CEVTCLPSHWTRLPDISWELGLLVSHSSY  162
         |     ||||||:|||            |||||||||||||||:||:||||||||||
Db  231 EPVPLEEVQLVV---EPEGGAVAPGGTVTLTCEVPAQPS---PQIHWMKD-------  274

Qy  163 YFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINI  222
         ||||||||||||||||||||||||||||||||||||||||||:||||||||:||||||||
Db  275 -GVPLPLPSPVLLPEIGPQDQGTYSCVAHSHSGPQESRAVSISIIE-PGEEG-------  327

Qy  223 PGVLSSLPSLGFSLPTWGKVGLGLLAGTMLLI--------------PTCTLTLRCCCCRRRCCGCNCC  275
         |||||||||||||||||| |||:|| ||:|||              |||||||||||||||| |||
Db  328 -------------PTAGSVGGSGLGTLALALGILGGIGTAALLIGVILWQRR------------  366

Qy  276 CRCCFCCRRKRGFRIQPFQKKSEKEKI--NKETETESGNENSGYNSDEQKTDTRASLPPKS  333
                                     ||:||||:|||:|||||||||:|||:||||||||:
Db  367 -------------------ORRGEERKAPENQEEEERAELN-------------------  389

Qy  334 CESSDPEQRNSSCGPP  349
         |::||||||||||||
Db  390 -QSEEPBAGESSTGGP  404
```

```
RESULT 1 (continued header)
Query Match       ;  Score 133;  DB 2;  Length 1427;
```

RESULT 5

A54100

Db   237 QKGTLQCEASAVPSAEFQ---WFK 257
              |:|| | ||
tumor suppressor protein DCC precursor - human
N;Alternate names: colorectal cancer suppressor DCC
C;Species: Homo sapiens (man)
C;Date: 02-Aug-1994 #sequence_revision 02-Aug-1994 #text_change 05-Nov-1999
A;Reference number: A54100; MUID:95011532; PMID:7926722
A;Accession: A54100
A;Molecule type: mRNA
R;Hedrick, L.; Cho, K.R.; Fearon, E.R.; Wu, T.C.; Kinzler, K.W.; Vogelstein, B.
Genes Dev. 8, 1174-1183, 1994
A;Title: The DCC gene product in cellular differentiation and colorectal tumorigenesis.
A;Reference number: A54100; MUID:95011532; PMID:7926722
A;Accession: A54100
A;Molecule type: mRNA
A;Residues: 1-1447 <HED>
A;Cross-references: EMBL:X76132; NID:g453209; PIDN:CAA53735.1; PID:g453210
R;Fearon, E.R.; Cho, K.R.; Nigro, J.M.; Kern, S.E.; Simons, J.W.; Ruppert, J.M.; Hamilt
Science 247, 49-56, 1990
A;Title: Identification of a chromosome 18q gene that is altered in colorectal cancers.
A;Reference number: A40098; MUID:90100559; PMID:2294591
A;Accession: A40098
A;Molecule type: mRNA
A;Residues: 1-750 <FEA>
A;Cross-references: GB:M32292; NID:g181492; PIDN:AAA35751.1; PID:g181493
A;Gene: GDB:DCC
A;Genetics:
A;Cross-references: GDB:l19838; OMIM:120470
A;Map position: 18q21.1-18q21.1
A;Keywords: transmembrane protein; tumor suppressor
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1447/Product: tumor suppressor protein DCC precursor #status predicted <MAT>

Query Match   6.2%; Score 128; DB 2; Length 1447;
Best Local Similarity   24.8%;
Matches   60; Conservative 33; Mismatches 99; Indels 50; Gaps 10;

Qy   10 SGSGNEV----------------IEGPQNATVLKGSQARFNCTVSQGW--KLIMWALSDM 51
              | ||| | ::| | ||| | :|:| |
Db   220 SRYGNEAEVRILSDPGLHRQLYFLQRSPSNVVAIEGKDAVLECCVS-GYPPSFTWLRGEE 278

Qy   52 VVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSA 109
              | | | : | ||| :| || || : | :|| : | | ||
Db   279 VI---------------QLRSKKYSLLGG--SNLLISNVTDDDSGMYYCVVTYKNENISASA 323

Qy   110 YLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPFDISW-ELGLLVSHSSYYFVPEP 168
              ||| | | | :| | | | | :||| |
Db   324 ELTVLVPPWFLNHPSNLYAYESMDIEFECTVS-GKPVPTVNWMKNGDVVIPSDYFQIVGG 382

Qy   169 SDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSS 228
              ||| | :| | :|||
Db   383 SNLR------ILGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS------VLPS 430

Qy   229 LP 230

Db   431 AP 432

RESULT 6

I48696

protein-tyrosine kinase (EC 2.7.1.112) nsk2 precursor, splice form 2 - mouse
N;Alternate names: receptor-type tyrosine kinase
N;Contains: protein-tyrosine kinase nsk2 precursor, splice form 4
C;Species: Mus musculus (house mouse)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C;Accession: I48696; S60738
R;Ganju, P.; Walls, E.; Brennan, J.; Reith, A.D.
Oncogene 11, 281-290, 1995
A;Title: Cloning and developmental expression of Nsk2, a novel receptor tyrosine kinase
A;Reference number: I48696; MUID:95349951; PMID:7624144
A;Accession: I48696

A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-871 <GAN1>
A;Cross-references: EMBL:X86444; NID:g929723; PIDN:CAA60165.1; PID:g929724
A;Molecule type: DNA
A;Residues: 1-456,'A',466-871 <GAN2>
A;Cross-references: EMBL:X86444; NID:g929723
A;Experimental source: splice form 4
A;Experimental source: splice form 2
C;Comment: For alternate splice forms see PIR:I48697.
C;Genetics:
A;Gene: nsk2
A;Cross-references: MGI:103308
A;Superfamily: mouse ror-related receptor; immunoglobulin homology; protein kinase ho
C;Keywords: ATP; glycoprotein; phosphotransferase; receptor; transmembrane protein; t
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-871/Product: protein-tyrosine kinase nsk2, splice form 2 #status predicted <MAT2
F;22-456,'A',466-871/Product: protein-tyrosine kinase nsk2, splice form 4 #status pre
F;42-101/Domain: immunoglobulin homology <IMM1>
F;135-192/Domain: immunoglobulin homology <IMM2>
F;226-284/Domain: immunoglobulin homology <IMM3>
F;498-518/Domain: transmembrane #status predicted <TRM>
F;575-865/Domain: protein kinase homology <KIN>
F;583-591/Region: protein kinase ATP-binding motif
F;222,462/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   6.2%; Score 127.5; Pred, No. 0.13; DB 1; Length 871;
Best Local Similarity   21.7%;
Matches   70; Conservative 41; Mismatches 125; Indels 87; Gaps 16;

Qy   15 EVIEGPQNATVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY 73
              | | ||| ||| | | ||: ||| | : | | ::|
Db   122 KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD---------NALRENSRIAALE- 171

Qy   74 DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-------LTVQVMGELFIPSVNLVV 128
              | ||| || || | | | | | :|
Db   172 -----SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKLVKLEVLVGRILRAPESHNV 223

Qy   129 AENEPCEVTC----LPSHWTRLPFDISW--ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQ 183
              | | | :||| |
Db   224 TFGSFVTLRCTEIGIP---VPTISWIENGNAVSSGSIQESVKDRVIDSRLQLFITKP- 277

Qy   184 SNGTLTCVAT--WKSLKARKSATVNLTVIR----CPQDTGGGINIPG-----V 225
              | |: | :
Db   278 --GLYTCIATNKHGEKFSTAKAAATVSIAEWSKSQKDSQGYCAQYRGEGVLMQGPGEKML 335

Qy   226 LSSLPSLGFSLP----TWGKVGLGLAGTMLLTPTCLTIRCCCRRRCCGCN-- 273

Db   336 LVFLPFTSHRDPEDAQELLIHTAWNEL------KAVSPICRPAAEALLCYLHLFLECSPG 388

Qy   274 ------CCCRCC-----FCCR 283

Db   389 VVPTPMPICREYCLAVKELFCAK 411

RESULT 7

I48697

protein-tyrosine kinase (EC 2.7.1.112) nsk2 precursor, splice form 1 - mouse
N;Alternate names: receptor-type tyrosine kinase
N;Contains: protein-tyrosine kinase nsk2 precursor, splice form 3
C;Species: Mus musculus (house mouse)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C;Accession: I48697; S60740
R;Ganju, P.; Walls, E.; Brennan, J.; Reith, A.D.
Oncogene 11, 281-290, 1995
A;Title: Cloning and developmental expression of Nsk2, a novel receptor tyrosine kina
A;Reference number: I48696; MUID:95349951; PMID:7624144
A;Accession: I48697

A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-881 <GAN1>
A;Cross-references: EMBL:X86445; NID:g929725; PIDN:CAA60166.1; PID:g929726

A;Experimental source: splice form 1
A;Accession: S60740
A;Molecule type: DNA
A;Residues: 1-456, 'A', 466-881
A;Cross-references: EMBL:X86445; NID:g929725 <GAN2>
A;Experimental source: splice form 3
C;Comment: For alternate splice forms see PIR:I48696.
C;Genetics:
A;Gene: nsk2
A;Cross-references: MGI:103308
C;Superfamily: mouse ror-related receptor; immunoglobulin homology; protein kinase homol
C;Keywords: ATP; glycoprotein; phosphotransferase; receptor; transmembrane protein; tyr
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-881/Product: protein-tyrosine kinase nsk2, splice form 3 #status predic
F;22-456,'A',466-881/Product: protein-tyrosine kinase nsk2, splice form 1 #status predicted <MAT1>
F;42-101/Domain: immunoglobulin homology <IMM1>
F;135-192/Domain: immunoglobulin homology <IMM2>
F;226-284/Domain: immunoglobulin homology <IMM3>
F;498-518/Domain: transmembrane #status predicted <TRM>
F;575-865/Domain: protein kinase homology <KIN>
F;583-591/Region: protein kinase ATP-binding motif
F;222,462/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   6.2%;   Best Local Similarity   21.7%;
Matches   70;   Conservative   41;   Mismatches 125;   Indels   87;   Gaps   16;
Score 127.5;   DB 1;   Length 881;

```
Qy    15 EVIEGPQNATVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY 73
Db   122 KIRPPINVKILEGLKAVLPCTTMGNPKPSVSWIKGD------NALRENSRIAALE- 171

Qy    74 DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY----LIVQVMGELFIPSVNLVV 128
Db   172 --SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKIVLKLEVLGRILRAPESHNV 223

Qy   129 AENEPCEVTC------LPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQ 183
Db   224 TFGSEVTLRCTEIGIP-----VPIISWIENGNAVSSGSIQESVKDRVIDSRLQLFITKP- 277

Qy   184 SNGTLTCVAT------WKSLRARKSATVNLTVIR------CPQDTGGGINIPG-----V 225
Db   278 --GLYTCIAINKHGEKFSTAKAAATVSIAEWSKSQKDSQGYCAQYRGEGVLMQGPGEKML 335

Qy   226 LSSLPSLGFSLP---------TWGSKVGLGLAGTMLLTPTCLTIRCCCCRRRCCGCN-- 273
Db   336 LVFLPTTSHRDPEDAQELLIHTAWNEL------KAVSPLCRPAAAFALLCVHLFLECSPG 388

Qy   274 ------CCCRCC------FCCR 283
Db   389 VVPTPMPICREYCLAVKELFCAK 411
```

RESULT 8
T29757
protein UNC-89 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C;Accession: T29757
R;Du, Z.; Le, T.T.; Wilson, R.
R; submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid C09D1.
A;Reference number: Z20679
A;Accession: T29757
A:Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6642 <DUZ>
A;Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
A;Experimental source: strain Bristol N2; clone C09D1
C;Genetics:
A;Gene: CESP:unc-89
A;Map position: 1
A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

RESULT 9
JC4593
protein-tyrosine kinase-related receptor PTK7 precursor - human
N;Alternate names: receptor protein tyrosine kinase-like protein (RPTK)
C;Species: Homo sapiens (man)
C;Date: 16-Apr-1996 #sequence_revision 24-May-1996 #text_change 24-Sep-1999
C;Accession: JC4593
R;Park, S.K.; Lee, H.S.; Lee, S.T.
J; Biochem. 119, 235-239, 1996
A;Title: Characterization of the human full-length PTK7 cDNA encoding a receptor
A;Reference number: JC4593; MUID:97037064; PMID:8882711
A;Accession: JC4593
A;Molecule type: mRNA
A;Residues: 1-1070 <PAR>
A;Cross-references: GB:U40271; NID:g1322231; PIDN:AAC50484.1; PID:g1322232
C;Comment: This protein is a member of receptor protein tyrosine kinase family, but p
C;Genetics:
A;Gene: GDB:PTK7
A;Cross-references: GDB:1134760; OMIM:601890
A;Map position: 6p21.1-6p12.2
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C;Keywords: cell adhesion; extracellular protein; glycoprotein; phosphotransferase; t
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1070/Product: protein-tyrosine kinase 7 #status predicted <MAT>
F;31-703/Domain: extracellular #status predicted <EXC>
F;704-725/Domain: transmembrane #status predicted <TMM>
F;726-1070/Domain: intracellular #status predicted <INC>
F;794-1065/Domain: protein kinase homology <KIN>
F;116,175,184,214,268,283,405,463,567,646/Binding site: carbohydrate (Asn) (covalent)

Query Match   6.18;   Best Local Similarity   23.9%;
Matches   52;   Conservative   39;   Mismatches   92;   Indels   35;   Gaps   8;
Score 125.5;   DB 2;   Length 6642;

```
Qy    16 VIEGPQNATVLKGSQARFNCTVSQGW-KLIMWALSDMVVLSVRPMEPIITNDRFTSQRY 73
Db  2077 VVDGPKSVTIKETETAEFKADIS-GFPAPTVKWIINEKIVEESRTIITTIKTEDVYT-- 2131

Qy    74 DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLIVQVMGELFIPS------ 128
Db  2132 --------LKISNAKIEQTGTVKVTAQNSAGQDSKQADLKVEPNVRAPKFRKSQLTDKVA 2182

Qy   129 AENEPCEVTCL---PSHWTRLPDISWEL-GLLVSHSSYYFVPEPSDLQSAVSILALTPQS 184
Db  2183 DEGEPLRWNLEDLDGPSPGT---EVSWLLNGQPLIKSDTVQQVDHGDGTYHVTIAEAKPEM 2239

Qy   185 NGTLTCVATWKSLRARKSATVNLTV----------VIRCPQD 215
Db  2240 SGTLIFAKAKNAAGECETSAKVIVNGGNKKPEFVQAPQN 2277
```

Query Match   6.0%;   Best Local Similarity   25.3%;
Matches   58;   Conservative   32;   Mismatches 103;   Indels   36;   Gaps   10;
Score 124;   DB 2;   Length 1070;

```
Qy    16 VIEGPQNATVLKGSQARFNCTVS-QGWKLIMWALSDMVVLSVRPMEPIITNDRFTSQRYD 74
Db   227 VVLAPQDVVARYEEAMFHCQFSAQPPPSLQWLFEDETPITNRSPPHLRRATVFA---- 282

Qy    75 QGGNFTSEMIIHNVEPSDSGNIRCSLQNSR-----LHGSAYLIVQVMGELFIPSVNLVVA 129
Db   283 -----NGSLLITQVRPRNAGIYRC1GQGQRGPPIILEATLHLAEIEDMPLFEPRVFTAGS 337

Qy   130 ENEPCEVTCLPSHWTRLPDISWE-LGL-LVSHSSYYFVPEPSDLQSAVSILALTPQSN-G 186
Db   338 EE---RVTCLPPKGLPEPSVWWEHAGVRLPTHGRVV------QKGHELVLANIAESDAG 387

Qy   187 TLTCVATWKSLKARKSATVNLTV------IRCPQDTGGGINIPGVLSSL 229
Db   388 VYTCHAA--NLAGQRRQDVNITVATVPSWLKPQDSQLEEGKPGYLDCL 434
```

RESULT 10
T15651
hypothetical protein C27A2.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15651
R;Nhan, M.
submitted to the EMBL Data Library, May 1996
A;Description: The sequence of C. elegans cosmid C27A2.
C;Accession: Z18382
A;Reference number: T15651
A;Accession: T15651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-188 <NHA>
A;Cross-references: EMBL:U58760; NID:g1330384; PID:g1330389; PIDN:AAB00710.1; GSPDB:GN00
A;Experimental source: strain Bristol N2; clone C27A2
C;Genetics:
A;Gene: CESP:C27A2.5
A;Map position: 2
A;Introns: 19/3; 91/2

Query Match 5.9%; Score 122.5; DB 2; Length 188;
Best Local Similarity 61.3%; Pred. No. 0.059;
Matches 19; Conservative 1; Mismatches 8; Indels 3; Gaps 2;

```
Qy 256 CTLTIRCCCCRRRCCGG--NCCC-RCCFCCR 283
           |||| |||| ||| | |||| |||| |||
Db  86 CCCRPKCCCCRRCCTCCRTCCCTRCCTCCR 116
```

RESULT 11
JC4025
opioid-binding cell adhesion protein - human
C;Species: Homo sapiens (man)
C;Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 17-Mar-2000
C;Accession: JC4025
R;Shark, K.B.; Lee, N.M.
Gene 155, 213-317, 1995
A;Title: Cloning, sequencing and localization to chromosome 11 of a cDNA encoding a huma
A;Reference number: JC4025; MUID:95237612; PMID:7721093
A;Accession: JC4025
A;Molecule type: mRNA
A;Residues: 1-345 <SHA>
A;Cross-references: GB:L34774; NID:g514373; PIDN:AAA36387.1; PID:g514374
A;Experimental source: brain
C;Comment: This protein binds opioid alkaloids in the presence of acidic lipids, exhibit
C;Genetics:
A;Gene: GDB:OPCML; OBCAM; OPCM
A;Cross-references: GDB:251677; OMIM:600632
A;Map position: 11pter-11qter
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin

Query Match 5.8%; Score 120; DB 2; Length 345;
Best Local Similarity 25.0%; Pred. No. 0.17;
Matches 71; Conservative 37; Mismatches 102; Indels 74; Gaps 18;

```
Qy   4 VIFLHGSG----SGNEVI-EGPQNATVLKGSQARFNCIVSQGWKLIMWALSDMVVL---- 54
        :||| :|          ::|  | ||| |  || | |  ||| ||:
Db  21 LLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAGN 79

Qy  55 --SVRPMEPIITNDRFTSQRYDQQGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRL 105
          ::|  |  | || |:|  | | || :|: |   :||| |:| |  :|
Db  80 DKWSIDPRVILVN---TPTQY-----SIMIQNVDVYDEGPYTCSVQTDNHPKTSRV 128

Qy 106 HGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCL----PSHWTRLPDISWELGLLVSHSS 161
        | |  :|:  :|:  |  |: :|  : || | |       |||| | :|:||| |||
Db 129 H----LIVQVPPQIMNISSDVTVNEGSSVTLLCLAIGRPE----PTVTWR-----HLS 173

Qy 162 YY-----FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-----IRC 212
        ::     |  |  :::|  || :|    || | :| |   | | |:: :|| |      | :
Db 174 VKEGQGFVSEDEYLE----ISDIKRDQSGEYECSAL-NDVAAPDVRKVKITVNYPPYISK 228

Qy 213 PQDTGGGINIPGVL----SSLPSLGFSLPTWGK---VGLGLAG 248
        | :|||:|||| |      ::|  |  | | ||     | | |
Db 229 AKNTGVSVGQKGILSCEASAVPMAEFQ---WFKEETRLATGLDG 269
```

RESULT 12
S03199
opioid-binding protein OPCAM precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 17-Mar-2000
C;Accession: S03199
R;Schofield, P.R.; McFarland, K.C.; Hayflick, J.S.; Wilcox, J.N.; Cho, T.M.; Roy, S.;
EMBO J. 8, 489-495, 1989
A;Title: Molecular characterization of a new immunoglobulin superfamily protein with
A;Reference number: S03199; MUID:89251576; PMID:2721489
A;Accession: S03199
A;Molecule type: mRNA
A;Residues: 1-345 <SCH>
A;Cross-references: EMBL:X12672; NID:g585; PIDN:CAA31192.1; PID:g586
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-ter
C;Keywords: transmembrane protein
A;Molecule type: signal sequence
F;1-27/Domain: signal sequence
F;28-345/Product: opioid-binding protein OPCAM #status predicted <SIG>
#status predicted <MAT>

Query Match 5.8%; Score 119; DB 2; Length 345;
Best Local Similarity 25.1%; Pred. No. 0.2;
Matches 70; Conservative 36; Mismatches 103; Indels 70; Gaps 17;

```
Qy   4 VIFLHGSG----SGNEVI-EGPQNATVLKGSQARFNCIVSQGWKLIMWALSDMVVL---- 54
        :||| :|          ::|  | ||| |  || | |  ||| ||:
Db  21 LLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAGN 79

Qy  55 --SVRPMEPIITNDRFTSQRYDQQGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRL 105
          ::|  |  | || |:|  | | || :|: |   :||| |:| |  :|
Db  80 DKWSIDPRVILVN---TPTQY-----SIMIQNVDVYDEGPYTCSVQTDNHPKTSRV 128

Qy 106 HGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCL----PSHWTRLPDISWELGLLVSHSS 161
        | |  :|:  :|:  |  |: :|  : || | |       |||| | :|:||| |||
Db 129 H----LIVQVPPQIMNISSDVTVNEGSSVTLLCLAIGRPE----PTVTWR-----HLS 173

Qy 162 YY-----FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-----IRC 212
        ::     |  |  :::|  || :|    || | :| |   | | |:: :|| |      | :
Db 174 VKEGQGFVSEDEYLE----ISDIKRDQSGEYECSAL-NDVAAPDVRKVKITVNYPPYISK 228

Qy 213 PQDTGGGINIPGVL----SSLPSLGFSLPTWGKVGLGLA 247
        | :|||:|||| |      ::|  |  | | ||  | | |
Db 229 AKNTGVSVGQKGILSCEASAVPMAEFQ---WFKEDTRLA 264
```

RESULT 13
JC1239
opioid-binding protein (clones SG8 and SG13) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 19-May-2000
C;Accession: JC1239
R;Lippman, D.A.; Lee, N.M.; Loh, H.H.
Gene 117, 249-254, 1992
A;Title: Opioid-binding cell adhesion molecule (OBCAM)-related clones from a rat brai
A;Reference number: JC1238; MUID:93477701; PMID:1339369
A;Accession: JC1239
A;Molecule type: mRNA
A;Residues: 1-345 <LIP>
A;Cross-references: GB:M88710; NID:g2032247; PIDN:AAA40859.1; PID:g2032248; GB:M88711;
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-ter
C;Keywords: transmembrane protein

Query Match 5.7%; Score 118; DB 2; Length 345;
Best Local Similarity 24.7%; Pred. No. 0.24;
Matches 71; Conservative 38; Mismatches 104; Indels 74; Gaps 18;

```
Qy  4 VIFLHGSG----SGNEVI-EGPQNATVLKGSQARFNCIVSQGWKLIMWALSDMVVL---- 54
       :||| :|          ::|  | ||| |  || | |  ||| ||:
Db 21 LLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAGN 79
```

```
Qy   55  ---SVRPMEPIITNDRFTSQRYDGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRL 105
            :|| ||| |||  |: |        :|| ||    || |:|
Db   80  DKWSIDPRVILVN---TPTQY-----SIMIQNVDVDEGPYTCSVQTDNHPKTSRV 128

Qy  106  HGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCL-----PSHWTRLPDISWELGLLVSHSS 161
            ::|            |: :|  ::|   ||          :||
Db  129  H----LIVQVPPQIMNISSDITVNEISSVTLLCLAIGRPE----PTVTWR----HLS 173

Qy  162  YY----FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-----IRC 212
            ||         :|  |||            ||
Db  174  VKEGQGFVSEDEYLE----ISDIKRDQSGEYECSAL-NDVAAPDVRKVKIITVNYPPYISK 228

Qy  213  PQDTGGGINIPGVL----SSLPSLGFSLPTWGK-----VGLGLAGTML 251
            |||            ||              |||  || |
Db  229  AKNTGVSVGQKGILSCEASAVPMAEFQ----WFKEDTRLATGLDGVRI 272
```

RESULT 14
I50600
neogenin - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I50600
R;Vielmetter, J.; Kayyem, J.F.; Roman, J.M.; Dreyer, W.J.
J. Cell Biol. 127, 2009-2020, 1994
A;Title: Neogenin, an avian cell surface protein expressed during terminal neuronal diff
A;Reference number: A55193; MUID:95105243; PMID:7806578
A;Accession: I50600
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1443 <VIE>
A;Cross-references: EMBL:U07644; NID:g641965; PIDN:AAC59662.1; PID:g641966

Query Match   5.7%; Score 117.5; DB 2; Length 1443;
Best Local Similarity 26.2%; Pred. No. 1.2;
Matches   49; Conservative   28; Mismatches   73; Indels   37; Gaps   11;

```
Qy   17  IEGPQNATVLKGSQARFNCTVSQGW----KLIMWALSDMVVLSVRPMEPIITNDRFTSQY- 73
            |||                        |:   ||
Db  218  VRQPSSLTKVTGQNAVPPC-VAGGFPPYVRWTKNG-------EELITED---SERFA 264

Qy   74  -DQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAE 130
            |:|||                        :|| ||
Db  265  LRAGGS----LLISDVTEEDVGTYTCIADNENETIEAQAELAVQVPPEFLKRPANIYAHE 320

Qy  131  NE-----PCEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSN 185
            ||||  ::|  :|  :| :| ::|
Db  321  SMDIVFECEVTGKPI----PTVKWVKNGDVVIPSDYFKIVKEHNLQ---VLGLVKSDE 371

Qy  186  GTLTCVA 192
            |:| |:|
Db  372  GFYQCIA 378
```

Query Match   5.7%; Score 116.5; DB 2; Length 338;
Best Local Similarity 25.0%; Pred. No. 0.31;
Matches   66; Conservative   33; Mismatches   96; Indels   69; Gaps   16;

```
Qy   22  NATVLKGSQARFNCTVSQGWKLIMWALSDMVVL-------SVRPMEPIITNDRFTSQRYD 74
            |  |  || ||| :|                         |||
Db  186  NVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAGNDKWSIDPRVILVN---TPTQY- 91

Qy   75  QGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRLHGSAYLTVQVMGELFIPSVNLVV 128
            ::|            |||| |                 :||||
Db   92  --SIMIQNVDVDEGPYTCSVQTDNHPKTSRVH---LIVQVPPQIMNISSDITV 140

Qy  129  AENEPCEVTCL-----PSHWTRLPDISWELGLLVSHSSYY----FVPEPSDLQSAVSILAL 180
            :|   ||          :||                        ||
Db  141  NEISSVTLLCLAIGRPE----PTVTWR----HLSVKEGQGFVSEDEYLE---ISDI 185

Qy  181  TPQSNGTLTCVATWKSLKARKSATVNLTV-----IRCPQDTGGGINIPGVL----SSLPS 231
                         ||                    ||| |              ||:||
Db  186  KRDQSGEYECSAL-NDVAAPDVRKVKITVNYPPYISKAKNTGVSVGQKGILSCEASAVPM 244

Qy  232  LGFSLPTWGK---VGLGLAGTML 251
            |   |||  || || |
Db  245  AEFQ----WFKEDTRLATGLDGVRI 265
```

Search completed: April 28, 2003, 21:08:48
Job time : 26.8683 secs

RESULT 15
JC1238
opioid-binding protein (clone DUZ1) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 19-May-2000
C;Accession: JC1238
R;Lippman, D.A.; Lee, N.M.; Loh, H.H.
Gene 117, 249-254, 1992
A;Title: Opioid-binding cell adhesion molecule (OBCAM)-related clones from a rat brain c
A;Reference number: JC1238; MUID:92347701; PMID:1339369
A;Accession: JC1238
A;Molecule type: mRNA
A;Residues: 1-338 <LIP>
A;Cross-references: GB:M88709; NID:g203245; PIDN:AAA40858.1; PID:g203246
A;Experimental source: brain
A;Genetics:
A;Gene: OBCAM
C;Gene: OBCAM
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin

OM protein - protein search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.4_p5_4578

Run on: April 29, 2003, 11:05:56 ; Search time 18.8683 seconds
(without alignments)
1622.277 Million cell updates/sec

Title: US-09-729-264-2
Perfect score: 2059
Sequence: 1 MGLVIFLHSGSGNEVIEGP..........HPQASFNLASPEKVSNTTVV 382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues
Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published_Applications_AA:*
                1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
               10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
               11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
               12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
               13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
               14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	142	6.9	405	1	US-08-755-235-4 Sequence 4, Appli
2	138.5	6.7	2473	9	US-10-184-644-559 Sequence 559, App
3	138.5	6.7	2473	9	US-10-184-634-559 Sequence 559, App
4	136	6.6	708	9	US-10-174-590-584 Sequence 584, App
5	136	6.6	708	9	US-10-175-737-584 Sequence 584, App
6	136	6.6	708	9	US-10-176-758-584 Sequence 584, App
7	136	6.6	708	9	US-10-173-706-584 Sequence 584, App
8	136	6.6	708	9	US-10-175-738-584 Sequence 584, App
9	136	6.6	708	9	US-10-175-752-584 Sequence 584, App
10	136	6.6	708	9	US-10-176-482-584 Sequence 584, App
11	136	6.6	708	9	US-10-176-757-584 Sequence 584, App
12	136	6.6	708	9	US-10-176-913-584 Sequence 584, App
13	136	6.6	708	9	US-10-180-552-584 Sequence 584, App
14	136	6.6	708	9	US-10-180-557-584 Sequence 584, App
15	136	6.6	708	9	US-10-173-700-584 Sequence 584, App
16	136	6.6	708	9	US-10-174-572-584 Sequence 584, App
17	136	6.6	708	9	US-10-174-579-584 Sequence 584, App
18	136	6.6	708	9	US-10-174-582-584 Sequence 584, App
19	136	6.6	708	9	US-10-174-588-584 Sequence 584, App
20	136	6.6	708	9	US-10-175-739-584 Sequence 584, App
21	136	6.6	708	9	US-10-175-740-584 Sequence 584, App
22	136	6.6	708	9	US-10-175-743-584 Sequence 584, App
23	136	6.6	708	9	US-10-176-488-584 Sequence 584, App
24	136	6.6	708	9	US-10-176-492-584 Sequence 584, App
25	136	6.6	708	9	US-10-176-747-584 Sequence 584, App
26	136	6.6	708	9	US-10-176-750-584 Sequence 584, App
27	136	6.6	708	9	US-10-176-985-584 Sequence 584, App
28	136	6.6	708	9	US-10-176-987-584 Sequence 584, App
29	136	6.6	708	9	US-10-176-991-584 Sequence 584, App
30	136	6.6	708	9	US-10-176-992-584 Sequence 584, App
31	136	6.6	708	9	US-10-176-993-584 Sequence 584, App
32	136	6.6	708	9	US-10-173-695-584 Sequence 584, App
33	136	6.6	708	9	US-10-173-697-584 Sequence 584, App
34	136	6.6	708	9	US-10-173-705-584 Sequence 584, App
35	136	6.6	708	9	US-10-174-576-584 Sequence 584, App
36	136	6.6	708	9	US-10-174-585-584 Sequence 584, App
37	136	6.6	708	9	US-10-174-586-584 Sequence 584, App
38	136	6.6	708	9	US-10-174-747-584 Sequence 584, App
39	136	6.6	708	9	US-10-175-481-584 Sequence 584, App
40	136	6.6	708	9	US-10-176-485-584 Sequence 584, App
41	136	6.6	708	9	US-10-176-487-584 Sequence 584, App
42	136	6.6	708	9	US-10-176-493-584 Sequence 584, App
43	136	6.6	708	9	US-10-176-495-584 Sequence 584, App
44	136	6.6	708	9	US-10-176-756-584 Sequence 584, App
45	136	6.6	708	9	US-10-176-911-584 Sequence 584, App

ALIGNMENTS

RESULT 1
US-08-755-235-4
Sequence 4, Application US/08755235
Publication No. US20030059423A1
GENERAL INFORMATION:
  APPLICANT: Stern, David M.
  APPLICANT: Schmidt, Ann Marie
  APPLICANT: Wu, Jun
  TITLE OF INVENTION: METHOD FOR TREATING SYMPTOMS OF DIABETES
  FILE REFERENCE: 0575/50159
  CURRENT APPLICATION NUMBER: US/08/755,235
  CURRENT FILING DATE: 1996-11-22
  NUMBER OF SEQ ID NOS: 4
  SOFTWARE: Patentin version 3.1
SEQ ID NO 4
  LENGTH: 405
  TYPE: PRT
  ORGANISM: Human
US-08-755-235-4

Query Match   86; Conservative   40; Mismatches 113; Indels 138; Gaps 19;
Best Local Similarity   22.8%; Pred. No. 0.0048;
Matches   86; Score 142;   DB 1;   Length 405;

```
Qy     15 EVIEGPQNATVLKGSQARFNCTVSQG-----WKLIMWALSDMVVLSVRPMEPIITNDRFTS-  70
          |::|                          | :|| |::     |  ::      |  ||
Db    125 EIVDSASELTA--GVPNKVGTCVSEGSYPAGTLSWHLDG------------KPLVPNERGVSV 173

Qy     71 ------QRYDQGGNFT--SEMIIHNYEPSDSGNIR----CSLQNSRLHGSAYLIVQVMGELF 120
                |  | :|       | |||  |   ||    | ::|   |  |   || :|
Db    174 KEQTRRHPEIGLFTLQSELM---VTPARGDRPFTSCSFSPGLPRHRALRTAPIQPRVW 230

Qy    121 IP----SVNLVVAENEP-----------------CEVTCLPSHWTRLPDISWELGLLSVSHSSY 162
          |  |  | |||                        |                |  ||
Db    231 EEVPLEEVQLVV---EPEGGAVAPGGTVTLTCEVPAQPS-----POIHWMKD-------- 274

Qy    163 YFVPEPSDLQSAVSISILALTPQSNGTLTCVATWKSLKARKS-ATVNLTVIRCPQDTGGGIN 221
          |  ||| |       |  | |:  || :|| ||    | | |     |  ||
Db    275 -GVPLPLPPSPVLILPEIGPDQGTYSCVATHSSHGPQESRAVVSISIIF-PGEEG----  328

Qy    222 IPGVLSSLPSLGFSLPTWGKVSLGLAGTMLLT------------PTCTLTIRCCCCRRCCGCNC 274
          |  |||                  |  | | |             ||:|
Db    275
```

```
Db   329 ----------PTAGSVGGSGLGTLALALGTILGLGLGTAALLIGVILWQRR------- 367
Qy   275 CCRCCFCCRRKRGFRIQFQKKSEKEKT--NKEETESGNENSGYNSDEQKTDTASLPPK 332
                                                              :|||:|
Db   368 ----------QRRGERKAPENQEEEERAELN------------------ 390
Qy   333 SCESSDPEQRNSCGPP 349
          |||| :||
Db   391 --QSEEPEAGESSTGGP 405
```

RESULT 2
US-10-184-644-559
; Sequence 559, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 559
; LENGTH: 2473
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-559

Query Match 6.7%; Score 138.5; DB 9; Length 2473;
Best Local Similarity 29.5%; Pred. No. 0.071;
Matches 31; Conservative 3; Mismatches 36; Indels 35; Gaps 2;

```
Qy   179 ALTPQSNGTLTCVATWKSLKARKSATVNLIVIRCPQDTGGGINIPGVLSSLPSLGFSLPT 238
             |||| ||||  ||||||||
Db  2274 AATTGAAGTTTCAATTAAAAATTTAATATGTTTCC------------------------ 2307

Qy   239 WGKVGLGLAGTMLLPTCTLTIRCCCCR-RRCCGCNCCCRCCFCC 282
               |||                ||||||
Db  2308 ------ATTCCTCATCGCCCACCCCACCCCCGCCCCCCACCACC 2344
```

RESULT 3
US-10-184-634-559
; Sequence 559, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634

; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 559
; LENGTH: 2473
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-559

Query Match 6.7%; Score 138.5; DB 9; Length 2473;
Best Local Similarity 29.5%; Pred. No. 0.071;
Matches 31; Conservative 3; Mismatches 36; Indels 35; Gaps 2;

```
Qy   179 ALTPQSNGTLTCVATWKSLKARKSATVNLIVIRCPQDTGGGINIPGVLSSLPSLGFSLPT 238
             |||| ||||  ||||||||
Db  2274 AATTGAAGTTTCAATTAAAAATTTAATATGTTTCC------------------------ 2307

Qy   239 WGKVGLGLAGTMLLPTCTLTIRCCCCR-RRCCGCNCCCRCCFCC 282
               |||                ||||||
Db  2308 ------ATTCCTCATCGCCCACCCCACCCCCGCCCCCCACCACC 2344
```

RESULT 4
US-10-174-590-584
; Sequence 584, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-584

Query Match 6.8%; Score 136; DB 9; Length 708;
Best Local Similarity 23.8%; Pred. No. 0.026;
Matches 55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;

```
Qy     3 LVIFLHGSGSGNEVIECPQNATVLKGSQARPNCTVSQGWKLIMWALSDMVVLSVRPMEPI 62
           |:     |||  |  :|||    |:|||  :|   | ||
Db    13 LFCFRGRAGPSPHFLQOPEDLVVLLGEEARLPCAIGAVWGLVQWTKSGLALGGQR----- 67

Qy    63 ITNDRTSQRXDQGGNPTS---EMIIHNYEPSDSGNIRCSLQNSRLHG-SAYLTV------ 113
           :|   :   | |      |::   |   |  |  ||  |::  |
Db    68 ---DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPE 124

Qy   114 --QVMGELFIPSVNLIVAENEPCEVTCLPFSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
             ||:|     |||:|       |:||     :  |    :|     |||  |:||
Db   125 APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVILLDGATFHQTLLKE 179

Qy   168 --PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLIVIRCPQDT 216
             |:     |:  | |:|  ||  ||  | ::|   | ||:  |:||  |
Db   180 GTPGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAILSLQYPPEVT 230
```

RESULT 5

US-10-176-758-584
Sequence 584, Application US/10176758
Publication No. US20030008353A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Chen,Jian
APPLICANT: Desnoyers,Luc
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C104
CURRENT APPLICATION NUMBER: US/10/176,758
CURRENT FILING DATE: 2002-06-21
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
LENGTH: 708
TYPE: PRT
ORGANISM: Homo Sapien
US-10-176-758-584

Query Match 6.6%; Score 136; DB 9; Length 708;
Best Local Similarity 23.8%; Pred. No. 0.026;
Matches 55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;

```
Qy    3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI 62
          : ||     :|  ||     |:| | |||  :|: | |:|:| ||
Db   13 LFCFRGRAGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLAIGGQGR----- 67

Qy   63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV----- 113
          |:     | |    | | :   ||  | ||       |||  :|  | |:
Db   68 ---DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPE 124

Qy  114 --QVMGELFIPSVNLVVAAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
          |:| :|   |  |:|| |  |  | |||   | |  |  :|  |:| :|: :|
Db  125 APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKE 179

Qy  168 --PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPDT 216
          | :||         || | |:|||    |  || :|| ||| :|
Db  180 GTPGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 7
US-10-173-706-584
Sequence 584, Application US/10173706
Publication No. US20030222293A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Chen,Jian
APPLICANT: Desnoyers,Luc
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C7
CURRENT APPLICATION NUMBER: US/10/173,706
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
LENGTH: 708
TYPE: PRT
ORGANISM: Homo Sapien
US-10-173-706-584

Query Match 6.6%; Score 136; DB 9; Length 708;
Best Local Similarity 23.8%; Pred. No. 0.026;
Matches 55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;

```
Qy    3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI 62
          : ||     :|  ||     |:| | |||  :|: | |:|:| ||
Db   13 LFCFRGRAGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLAIGGQGR----- 67

Qy   63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV----- 113
          |:     | |    | | :   ||  | ||       |||  :|  | |:
Db   68 ---DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPE 124

Qy  114 --QVMGELFIPSVNLVVAAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
          |:| :|   |  |:|| |  |  | |||   | |  |  :|  |:| :|: :|
Db  125 APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKE 179

Qy  168 --PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPDT 216
          | :||         || | |:|||    |  || :|| ||| :|
Db  180 GTPGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 6
US-10-175-737-584
Sequence 584, Application US/10175737
Publication No. US20030013153A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Chen,Jian
APPLICANT: Desnoyers,Luc
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C50
CURRENT APPLICATION NUMBER: US/10/175,737
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
LENGTH: 708
TYPE: PRT
ORGANISM: Homo Sapien
US-10-175-737-584

Query Match 6.6%; Score 136; DB 9; Length 708;
Best Local Similarity 23.8%; Pred. No. 0.026;
Matches 55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;

```
Qy    3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI 62
          : ||     :|  ||     |:| | |||  :|: | |:|:| ||
Db   13 LFCFRGRAGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLAIGGQGR----- 67

Qy   63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV----- 113
          |:     | |    | | :   ||  | ||       |||  :|  | |:
Db   68 ---DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPE 124

Qy  114 --QVMGELFIPSVNLVVAAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
          |:| :|   |  |:|| |  |  | |||   | |  |  :|  |:| :|: :|
Db  125 APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKE 179

Qy  168 --PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPDT 216
          | :||         || | |:|||    |  || :|| ||| :|
Db  180 GTPGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

Db 180 GTPGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230

RESULT 8
US-10-175-738-584
; Sequence 584, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C45
; CURRENT APPLICATION NUMBER: US/10,175,738
; CURRENT FILING DATE: 2002-06-19
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-584

Query Match
Best Local Similarity   6.6%;   Score 136; DB 9; Length 708;
Matches   55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;
; Similarity 23.8%; Pred. No. 0.026;

```
Qy   3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI  62
Db  13 LFCFRGRAGPSPHFLQQPEDLVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR-----  67

Qy  63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV---- 113
Db  68 ---DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPE 124

Qy 114 --QVMGELFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
Db 125 APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKE 179

Qy 168 --PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 216
Db 180 GTPGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 9
US-10-175-752-584
; Sequence 584, Application US/10175752
; Publication No. US20030022295A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C60
; CURRENT APPLICATION NUMBER: US/10,175,752
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-752-584

Query Match
Best Local Similarity   6.6%;   Score 136; DB 9; Length 708;
Matches   55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;
; Similarity 23.8%; Pred. No. 0.026;

```
Qy   3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI  62
Db  13 LFCFRGRAGPSPHFLQQPEDLVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR-----  67

Qy  63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV---- 113
Db  68 ---DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPE 124

Qy 114 --QVMGELFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
Db 125 APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKE 179

Qy 168 --PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 216
Db 180 GTPGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 10
US-10-176-482-584
; Sequence 584, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C70
; CURRENT APPLICATION NUMBER: US/10,176,482
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-482-584

Query Match
Best Local Similarity   6.6%;   Score 136; DB 9; Length 708;
Matches   55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;
; Similarity 23.8%; Pred. No. 0.026;

```
Qy   3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI  62
Db  13 LFCFRGRAGPSPHFLQQPEDLVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR-----  67

Qy  63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV---- 113
Db  68 ---DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPE 124

Qy 114 --QVMGELFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
```

RESULT 11
US-10-176-757-584
Sequence 584, Application US/10176757
; Publication No. US20030022297A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C86
; CURRENT APPLICATION NUMBER: US/10/176,757
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-757-584

  Query Match   6.6%; Score 136; DB 9; Length 708;
  Best Local Similarity 23.8%; Pred. No. 0.026;
  Matches 55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;

Qy   3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI 62
           || || || | |||| |||||
Db   13 LFCFRGRAGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR----- 67

Qy   63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLIV----- 113
Db   68 ---DLPGWSRYWISGNAANQOHDLHIRPVELEDEASYECOATQAGLRSRPAQLHVLVPPE 124

Qy 114 --QVMGELFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
Db 125 APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKE 179

Qy 168 --PSDLQSAVSILALTPQSNGTLLTCVATWKSLKARKSATVNLTVIRCPQDT 216
Db 180 GTPGSVESTLTLPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230

RESULT 12
US-10-176-913-584
Sequence 584, Application US/10176913
; Publication No. US20030022298A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C66
; CURRENT APPLICATION NUMBER: US/10/176,913
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-913-584

  Query Match   6.6%; Score 136; DB 9; Length 708;
  Best Local Similarity 23.8%; Pred. No. 0.026;
  Matches 55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;

Qy   3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI 62
           || || || | |||| |||||
Db   13 LFCFRGRAGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR----- 67

Qy   63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLIV----- 113

RESULT 13
US-10-180-552-584
Sequence 584, Application US/10180552
; Publication No. US20030022300A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C153
; CURRENT APPLICATION NUMBER: US/10/180,552
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-180-552-584

  Query Match   6.6%; Score 136; DB 9; Length 708;
  Best Local Similarity 23.8%; Pred. No. 0.026;
  Matches 55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;

Qy   3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI 62
           || || || | |||| |||||
Db   13 LFCFRGRAGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR----- 67

Qy   63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLIV----- 113
Db   68 ---DLPGWSRYWISGNAANQOHDLHIRPVELEDEASYECOATQAGLRSRPAQLHVLVPPE 124

Qy 114 --QVMGELFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
Db 125 APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKE 179

Qy 168 --PSDLQSAVSILALTPQSNGTLLTCVATWKSLKARKSATVNLTVIRCPQDT 216
Db 180 GTPGSVESTLTLPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230

RESULT 15
US-10-173-700-584
Sequence 584, Application US/10173700
Publication No. US20030027262A1
GENERAL INFORMATION:
  APPLICANT: Baker,Kevin P.
  APPLICANT: Chen,Jian
  APPLICANT: Desnoyers,Luc
  APPLICANT: Goddard,Audrey
  APPLICANT: Godowski,Paul J.
  APPLICANT: Gurney,Austin L.
  APPLICANT: Pan,James

  APPLICANT: Smith,Victoria
  APPLICANT: Watanabe,Colin K.
  APPLICANT: Wood,William I.
  APPLICANT: Zhang,Zemin
  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
  TITLE OF INVENTION: ACIDS ENCODING THE SAME
  FILE REFERENCE: P3430R1C14
  CURRENT APPLICATION NUMBER: US/10/173,700
  CURRENT FILING DATE: 2002-06-17
  Prior Application removed - See File Wrapper or Palm
  NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
  LENGTH: 708
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-173-700-584

Query Match   6.6%;   Score 136; DB 9;   Length 708;
Best Local Similarity   23.8%; Pred. No. 0.026;
Matches 55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;

Qy   3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI 62
          |||:| ||:|| :||||| ||| ||| |: |||||| |::|| ||| ||: :|
Db   13 LFCFRGRAGPSPHFLQQPEDLVLLGEEARLPCALGAYWGLVQWTKSGIALGGQR----- 67

Qy   63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV----- 113
          | | :|| :| :| ||:| | |||||: || | | ||| ||| |
Db   68 --DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYECQATQOAGLRSRPAQLHVLVPPE 124

Qy   114 --QVMGELFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
           :||||||||||||||| ||||:||||||||||| |||||| :| |||||||| ||||
Db   125 APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 179

Qy   168 --PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 216
          : :||||||||||||||||||||||||||||||||||||||||||| |
Db   180 GTPGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230

RESULT 14
US-10-180-557-584
Sequence 584, Application US/10180557
Publication No. US20030022301A1
GENERAL INFORMATION:
  APPLICANT: Baker,Kevin P.
  APPLICANT: Chen,Jian
  APPLICANT: Desnoyers,Luc
  APPLICANT: Goddard,Audrey
  APPLICANT: Godowski,Paul J.
  APPLICANT: Gurney,Austin L.
  APPLICANT: Pan,James
  APPLICANT: Smith,Victoria
  APPLICANT: Watanabe,Colin K.
  APPLICANT: Wood,William I.
  APPLICANT: Zhang,Zemin
  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
  TITLE OF INVENTION: ACIDS ENCODING THE SAME
  FILE REFERENCE: P3430R1C147
  CURRENT APPLICATION NUMBER: US/10/180,557
  CURRENT FILING DATE: 2002-06-25
  Prior Application removed - See File Wrapper or Palm
  NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
  LENGTH: 708
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-180-557-584

Query Match   6.6%;   Score 136; DB 9;   Length 708;
Best Local Similarity   23.8%; Pred. No. 0.026;
Matches 55; Conservative 40; Mismatches 106; Indels 30; Gaps 10;

Qy   3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI 62
          |||:| ||:|| :||||| ||| ||| |: |||||| |::|| ||| ||: :|
Db   13 LFCFRGRAGPSPHFLQQPEDLVLLGEEARLPCALGAYWGLVQWTKSGIALGGQR----- 67

Qy   63 ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV----- 113
          | | :|| :| :| ||:| | |||||: || | | ||| ||| |
Db   68 --DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYECQATQOAGLRSRPAQLHVLVPPE 124

Qy   114 --QVMGELFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
           :||||||||||||||| ||||:||||||||||| |||||| :| |||||||| ||||
Db   125 APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 179

Qy   168 --PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 216
          : :||||||||||||||||||||||||||||||||||||||||||| |
Db   180 GTPGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230

Search completed: April 29, 2003, 20:16:58
Job time : 28.8683 secs

us-09-729-264-2.rai

OM protein - protein search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.4_p5_4578

Run on: April 28, 2003, 18:14:43 ; Search time 13.9029 Seconds
(without alignments)
808.430 Million cell updates/sec

Title: US-09-729-264-2
Perfect score: 2059
Sequence: 1 MGLVIFLHGSGSGNEVIEGP...........HPQASFNIASPEKVSNTTVV 382

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB	ID	Description
1	128.5	6.2	1345	2	US-08-977-767-3	Sequence 3, Appli
2	128	6.2	1447	4	US-09-041-886-25	Sequence 25, Appl
3	128	6.2	1447	5	PCT-US94-05277-2	Sequence 2, Appli
4	126.5	6.1	869	2	US-08-374-834-16	Sequence 16, Appl
5	126.5	6.1	869	2	US-08-644-271-29	Sequence 29, Appl
6	126.5	6.1	869	5	US-09-077-955-33	Sequence 33, Appl
7	123.5	6.0	332	4	US-09-062-365-1	Sequence 1, Appl
8	122.5	5.9	340	4	US-09-651-200-2	Sequence 2, Appli
9	122.5	5.9	441	4	US-09-651-200-4	Sequence 4, Appl
10	122.5	5.9	340	6	US-09-651-200-6	Sequence 6, Appli
11	121.5	5.9	534	4	US-09-651-200-24	Sequence 24, Appli
12	121.5	5.9	534	2	US-08-633-148-4	Sequence 4, Appl
13	118.5	5.8	318	2	US-08-633-148-2	Sequence 2, Appl
14	118.5	5.8	340	20	US-09-651-200-20	Sequence 20, Appl
15	117	5.7	1395	4	US-09-540-245A-15	Sequence 15, Appl
16	114	5.5	868	4	US-08-374-834-1	Sequence 1, Appl
17	112	5.4	868	2	US-08-644-271-1	Sequence 1, Appli
18	112	5.4	868	5	US-09-077-955-1	Sequence 1, Appl
19	110	5.3	689	4	US-09-499-964-1	Sequence 1, Appl
20	109.5	5.3	478	5	PCT-US95-08493-15	Sequence 15, Appl
21	109.5	5.3	860	5	PCT-US95-08493-19	Sequence 19, Appl
22	109.5	5.3	868	5	PCT-US95-08493-21	Sequence 21, Appl
23	107.5	5.2	362	5	PCT-US95-08493-1	Sequence 1, Appli
24	107	5.2	365	2	US-08-979-424-3	Sequence 3, Appli
25	109.5	5.3	365	5	US-09-272-496-2	Sequence 2, Appli
26	106.5	5.2	946	5	PCT-US95-08493-13	Sequence 13, Appl
27	106	5.1	319	1	US-08-597-495B-22	Sequence 22, Appl
28	106	5.1	319	4	US-09-068-051A-22	Sequence 22, Appl
29	106	5.1	319	4	US-09-336-536-67	Sequence 67, Appl
30	106	5.1	319	4	US-09-254-465A-6	Sequence 6, Appli
31	106	5.1	801	4	US-07-906-349A-6	Sequence 6, Appli
32	105	5.1	285	3	US-08-482-085B-20	Sequence 20, Appli
33	105	5.1	1297	4	US-09-540-245A-17	Sequence 17, Appl
34	105	5.1	1381	4	US-09-540-245A-16	Sequence 16, Appl
35	104.5	5.1	879	1	US-08-554-612C-1	Sequence 1, Appli
36	104	5.1	421	2	US-08-659-984A-1	Sequence 1, Appl
37	104	5.1	421	4	US-08-660-531-1	Sequence 1, Appli
38	104	5.1	444	2	US-08-659-984A-5	Sequence 5, Appli
39	104	5.1	444	4	US-08-660-531-5	Sequence 5, Appli
40	103.5	5.0	890	1	US-08-445-640-2	Sequence 2, Appli
41	103.5	5.0	890	3	US-08-170-558-2	Sequence 2, Appli
42	103.5	5.0	890	4	US-08-447-314-2	Sequence 2, Appli
43	103.5	5.0	890	3	US-08-445-461-2	Sequence 2, Appli
44	103.5	5.0	911	3	US-08-286-305A-1	Sequence 1, Appli
45	103.5	5.0	911	2	US-08-441-104A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-977-767-3
Sequence 3, Application US/08977767

GENERAL INFORMATION:
Patent No. 5972684

APPLICANT: Bandman, Olga
APPLICANT: Yue, Henry
APPLICANT: Greenwald, Sara
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08,977,767
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0423 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1345 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1532042
US-08-977-767-3

Query Match 6.2%; Score 128.5; DB 2; Length 1345;

Left column

Best Local Similarity 33.0%; Pred. No. 0.0095; Mismatches 41; Indels 33; Gaps 5;
Matches 37; Conservative 1;

```
Qy  186 GTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGI----NIPGVLSSLPSLGFSLPTWGK 241
            |||
Db  414 GTCTCTGT------------------GC-CGTGGGAAGGGTCAGAGCCCGTGGATGTGGA 455

Qy  242 VGLGLAGTMLLT-PTCTLITIRCCCCRRRCCGCNCCCRC----------CFCC 282
            |||      |||||||
Db  456 CGTGAAGAGGGTCCTTCTATGACCCCTTCCTGCCCCTCTGAGACTCAGCACC 507
```

RESULT 2
US-09-041-886-25
Sequence 25, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-25

Query Match 24.8%; Score 128; DB 4; Length 1447;
Best Local Similarity 6.2%; Pred. No. 0.012; Matches 60; Conservative 33; Mismatches 99; Indels 50; Gaps 10;

```
Qy   10 SGSGGNEV----------------------IEGPQNATVLKGSQARFNCTVSQGW--KLIMWALSDM 51
            : |||
Db  220 SRTGNEAVRILSDPGLHRQLYFLQRPSNVVAIEGKDAVLECCVS-GYPPPSFTWLRGEE 278

Qy   52 VVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSA 109
            |: :|:|
Db  279 VI------------------QLRSKKYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNENISASA 323

Qy  110 YLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEP 168
            |:|:|
Db  324 ELTVLVPPWFLNHPSNLYAYESMDIEFECTVS-GKPVPTVNWMKNGDVVIPSDYFQIVGG 382

Qy  169 SDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSS 228
            |:|:|
Db  383 SNLR------ILGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS------VLPS 430
```

Right column

```
Qy  229 LP 230
Db  431 AP 432
```

RESULT 3
PCT-US94-05277-2
Sequence 2, Application PC/TUS9405277
; GENERAL INFORMATION:
; APPLICANT: Bruskin, Arthur
; APPLICANT: Jarosz, David E.
; APPLICANT: Johnson, Karen
; APPLICANT: Kinzler, Kenneth W.
; APPLICANT: Vogelstein, Bert
; APPLICANT: Zabrecky, James R.
; TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05277
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.42709
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202.508.9100
; TELEFAX: 202.508.9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-05277-2

Query Match 24.8%; Score 128; DB 5; Length 1447;
Best Local Similarity 6.2%; Pred. No. 0.012; Matches 60; Conservative 33; Mismatches 99; Indels 50; Gaps 10;

```
Qy   10 SGSGGNEV----------------------IEGPQNATVLKGSQARFNCTVSQGW--KLIMWALSDM 51
            : |||
Db  220 SRTGNEAVRILSDPGLHRQLYFLQRPSNVVAIEGKDAVLECCVS-GYPPPSFTWLRGEE 278

Qy   52 VVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSA 109
            |: :|:|
Db  279 VI------------------QLRSKKYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNENISASA 323

Qy  110 YLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEP 168
            |:|:|
Db  324 ELTVLVPPWFLNHPSNLYAYESMDIEFECTVS-GKPVPTVNWMKNGDVVIPSDYFQIVGG 382

Qy  169 SDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSS 228
            |:|:|
Db  383 SNLR------ILGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS------VLPS 430

Qy  229 LP 230
Db  431 AP 432
```

RESULT 4
US-08-374-834-16
Sequence 16, Application US/08374834
Patent No. 5656473
GENERAL INFORMATION:
    APPLICANT: Valenzuela, et al.
    TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTOR
    NUMBER OF SEQUENCES: 17
    CORRESPONDENCE ADDRESS:
       ADDRESSEE: Regeneron Pharmaceuticals, Inc.
       STREET: 777 Old Saw Mill River Road
       CITY: Tarrytown
       STATE: New York
       COUNTRY: USA
       ZIP: 10591
    COMPUTER READABLE FORM:
       MEDIUM TYPE: Floppy disk
       COMPUTER: IBM PC compatible
       OPERATING SYSTEM: PC-DOS/MS-DOS
       SOFTWARE: PatentIn Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
       APPLICATION NUMBER: US/08/374,834
       FILING DATE: 19-JAN-1995
       CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
       APPLICATION NUMBER: US 08/095,658
       FILING DATE: 21-JUL-1993
    ATTORNEY/AGENT INFORMATION:
       NAME: Cobert, Robert J.
       REGISTRATION NUMBER: 36,108
       REFERENCE/DOCKET NUMBER: REG 190A
    TELECOMMUNICATION INFORMATION:
       TELEPHONE: (914) 345-7400
       TELEFAX: (914) 345-7721
    INFORMATION FOR SEQ ID NO: 16:
       SEQUENCE CHARACTERISTICS:
          LENGTH: 869 amino acids
          TYPE: amino acid
          STRANDEDNESS:
          TOPOLOGY: unknown
          MOLECULE TYPE: protein
    US-08-374-834-16

Query Match   6.1%;   Score 126.5; DB 1; Length 869;
Best Local Similarity   20.9%; Pred. No. 0.0078;
    Matches   86; Conservative   51; Mismatches 176; Indels 99; Gaps 19;

Qy   15 EVIEGPQNATVLKGSQOARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY 73
          :| |||| ||||:||||::| ::| |:||::| ||| | | :| | | ::|
Db   122 KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD---------SPLRENSRIAVLE- 171

Qy   74 DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVWGELFIPSVNLVVAENE 132
          : ||||| |:||||||| ||:|| |:
Db   172 ------SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKVVKLEVEVFARILRAPESHNV 223

Qy   133 P---------CEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQ 183
          | :| | | | :|:| :|| : | | | :| | | | | ||
Db   224 TFGSFVTLHCTATGIP-----VPTITWIENGNAVSSGSIQESVKDRVIDSRLQLFITKP- 277

Qy   184 SNGTLTCVAT----WKSLKARKSATVNLTVIRCPQDTGGG---------INIPGVLSSLP 230
          | | | | | || | | |::|:|::||||| | | |:| | | |
Db   278 --GLYICIATNKHGEKFSTAKAAATISIAEWSKPQKDNKGYCAQYRGEVCNAVLAKDALV 335

Qy   231 SLGFSL----------PTWGKVGLGLAGTMLLTPTCTLTITRCCCCRRRCCGCN----- 273
          |:|||: | | ||:: : :: | | | | ::|
Db   336 FLNTSYADPEEAQELLVHTAWNEL--KVVSPVCRPAAEALLCNHIFQECSPGVVP 388

Qy   274 ----CCCRCC------FCCRRKRGRFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKT 323
                  :||:| | :| ::|:|: | || | : ::: |: ||
Db   389 TPIPICREYCLAVKELFCAKE---WLVMEEKTHRGLYRSEMHLLSVPECSKLPSMHWDP 444

Qy   324 TDFASLPPKSCESSDPEQRNSSCGPPHORADORP----PRPASHPQASFNLA 371

Db   445 TACARLP----- -HLDYNKENLKTFPP--MTSSKFPSVDIPNLPSSSSSSPSVS 489

RESULT 5
US-08-644-271-29
Sequence 29, Application US/08644271
Patent No. 5814478
GENERAL INFORMATION:
    APPLICANT: Valenzuela, et al.
    TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
    TITLE OF INVENTION: AND LIGANDS
    NUMBER OF SEQUENCES: 32
    CORRESPONDENCE ADDRESS:
       ADDRESSEE: Regeneron Pharmaceuticals, Inc.
       STREET: 777 Old Saw Mill Road
       CITY: Tarrytown
       STATE: NY
       COUNTRY: USA
       ZIP: 10591
    COMPUTER READABLE FORM:
       MEDIUM TYPE: Diskette
       COMPUTER: IBM Compatible
       OPERATING SYSTEM: DOS
       SOFTWARE: FastSEQ Version 2.0
    CURRENT APPLICATION DATA:
       APPLICATION NUMBER: US/08/644,271
       FILING DATE: 10-MAY-1996
       CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
       APPLICATION NUMBER: USSN 60/008,657
       FILING DATE: 15-DEC-1995
    ATTORNEY/AGENT INFORMATION:
       NAME: Cobert, Robert J
       REGISTRATION NUMBER: 36,108
       REFERENCE/DOCKET NUMBER: REG 195A
    TELECOMMUNICATION INFORMATION:
       TELEPHONE: 914-345-7400
       TELEFAX: 914-345-7721
       TELEX:
    INFORMATION FOR SEQ ID NO: 29:
       SEQUENCE CHARACTERISTICS:
          LENGTH: 869 amino acids
          TYPE: amino acid
          STRANDEDNESS:
          TOPOLOGY: unknown
          MOLECULE TYPE: protein
    US-08-644-271-29

Query Match   6.1%;   Score 126.5; DB 2; Length 869;
Best Local Similarity   20.9%; Pred. No. 0.0078;
    Matches   86; Conservative   51; Mismatches 176; Indels 99; Gaps 19;

Qy   15 EVIEGPQNATVLKGSQOARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY 73
          :| |||| ||||:||||::| ::| |:||::| ||| | | :| | | ::|
Db   122 KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD---------SPLRENSRIAVLE- 171

Qy   74 DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVWGELFIPSVNLVVAENE 132
          : ||||| |:||||||| ||:|| |:
Db   172 ------SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKVVKLEVEVFARILRAPESHNV 223

Qy   133 P---------CEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQ 183
          | :| | | | :|:| :|| : | | | :| | | | | ||
Db   224 TFGSFVTLHCTATGIP-----VPTITWIENGNAVSSGSIQESVKDRVIDSRLQLFITKP- 277

Qy   184 SNGTLTCVAT----WKSLKARKSATVNLTVIRCPQDTGGG---------INIPGVLSSLP 230
          | | | | | || | | |::|:|::||||| | | |:| | | |
Db   278 --GLYICIATNKHGEKFSTAKAAATISIAEWSKPQKDNKGYCAQYRGEVCNAVLAKDALV 335

Qy   231 SLGFSL----------PTWGKVGLGLAGTMLLTPTCTLTITRCCCCRRRCCGCN----- 273
          |:|||: | | ||:: : :: | | | | ::|
Db   336 FLNTSYADPEEAQELLVHTAWNEL--KVVSPVCRPAAEALLCNHIFQECSPGVVP 388

RESULT 6
US-09-077-955-33
; Sequence 33, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-33

Query Match   Best Local Similarity 20.9%; Score 126.5; DB 4; Length 869;
Matches 86; Conservative 51; Mismatches 176; Indels 99; Gaps 19;

```
Qy  15 EVIEGPQNATVLKGSQARFNCTVSQGGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY  73
Db 122 KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD----------SPLRENSRIAVLE- 171

Qy  74 DQGGNFISEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LIVQVMGELFIPSVNLVVAENE  132
Db 172 -GGSLRIHNVQKEDAGQYRCVAKNSL-GTAYSKVKLEVFARILRAPESHNV          223

Qy 133 P-------CEVTCLPSHWTLRLPDISW-ELGLLVSHSSYFVPEPSDLQSAVSIIALTPQ   183
Db 224 TFGSFVTLHCTATGIP-----VPTITWIENGNAVSSGSTQESVKDRVIDSRLQLFITKP- 277

Qy 184 SNGTLTCVAT----WKSLKARKSAATVNLTVTIRCPQDTGGG-                  230
Db 278 -GLYTCIATNKHGEKFSTAKAAAATISIAEWSKPQKDNKGYCAQYRGEVCNAVLAKDALV 335

Qy 231 SLGFSL-------PTWGKVGLGLAGTMLLTPTCLTIRCCCCRRRCGGCN-----       273
Db 336 FLNTSYADPEEAQELLVHTAWNEL------KVVSPVCRPAAEALICNHIFQECSPGVVP  388

Qy 274 ----CCCRCC------FCCRRKRGFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKT 323
Db 389 TPIPICREYCLAVKELFCAKE----WLVMEEKTHRGLYRSEMHLLSVPECSKLPSMHWDP 444

Qy 324 TDTASLPPKSCESSDPEQRNSSCGPPHQRADQRP-----PRPASHPQASFNLA       371
Db 445 TACARLP-------HLDYNKENLKTFPP-MTSSKPSVDIPNLPSSSSSFSVS        489
```

RESULT 7
US-09-062-365-1
; Sequence 1, Application US/09062365
; Patent No. 6465422
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Ann Marie
; APPLICANT: Stern, David
; TITLE OF INVENTION: METHOD FOR INHIBITING TUMOR INVASION OR SPREADING IN A
; TITLE OF INVENTION: SUBJECT
; FILE REFERENCE: 55424
; CURRENT APPLICATION NUMBER: US/09/062,365
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Human
US-09-062-365-1

Query Match   Best Local Similarity 24.18%; Pred. No. 0.0038; Score 123.5; DB 4; Length 332;
Matches 65; Conservative 33; Mismatches 89; Indels 83; Gaps 14;

```
Qy  15 EVIEGPQNATVLKGSQARFNCTVSQG--WKLIMWALSDMVVLSVRPMEPIITNDRFTS-   70
Db 103 EIVDSASELTA--GVPNRVGTCVSEGSSYPAGTLSWHLDG--------KPLVPNEKGVSV 151

Qy  71 ----QRYDQGGNFT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGSAYLTVQVMGELF 120
Db 152 KEQTRRHPETGLFTLQSELM--VTPARGGDPRPTFSCSFSPGLPRHRALRTAPIQPRVW  208

Qy 121 IP----SVNLVVAENEP------------CEVTCLPSHWTLRLPDISWELGLLVSHSSY  162
Db 209 EPVPLEEVQLVV---EPEGGAVAPGGTVTLTCEVPAQPS-----PQIHWMKD------   252

Qy 163 YFVPEPSDLQSAVSIIALTPQSNGTLTCVATWKSLKARKSAATVNLTVTIRCPQDTGGGINI 222
Db 253 -GVPLPLPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIE-PGEEG-----  305

Qy 223 PGVLSSLPSLGFSLPTWGKVGLGLAGTMLL                              252
Db 306 -------------------------PTAGSVGGSGLGTLAL                   321
```

RESULT 8
US-09-651-200-2
; Sequence 2, Application US/09651200
; Patent No. 6429303
; GENERAL INFORMATION:
; APPLICANT: Green et al
; TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
; TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: 15966-562 (CURA-62)
; CURRENT APPLICATION NUMBER: US/09/651,200
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: US/09/651,200
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/172909
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/152383
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: 60/183578
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-651-200-2

Query Match   Best Local Similarity 5.9%; Pred. No. 0.0049; Score 122.5; DB 4; Length 340;
Matches 77; Conservative 50; Mismatches 128; Indels 109; Gaps 17;

```
Qy  10 SGSGNEVIEGPQNATV-LKGSQARFNCTVS--QGWKL----IMWALSDMVVLSVRPMEPI  62
Db  48 SPTGAVEVQVPEDPVVALVGTDATLHCSFSPEPGFSLTQLNLIWQLTDTKQLV------- 100

Qy  63 IINDRFTSQRYDGGNF-----------------------------TSEMIIHNVEPSDSGNIRCSLQNSRLHGS 108
```

```
Db  101  ---HSFTEGR-DQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFV-SIRDFGS  155
Qy  109  AYLTVQVMGELFIPSVNLV----VAENEPCEVTCLPSHWTRLP--DISWELGL---LVSH  159
Db  156  AAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPLTGN  213
Qy  160  SSYYFVPEPSDLQSAVSIILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP---QDT  216
Db  214  VTTSQMANEQGLFDVHSVLRVVLGANGTYSC------------------------------  255
Qy  217  GGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCRRRCCGCNCCC  276
Db  256  HGSVTIIGQPMTFPPEAL-----WVTVGLSVCLIALLV------------------------  288
Qy  277  RCCFCCRRKRGFRIQFQKKSEKKTNKETESGNENSGYNSDEQKTTDTASLPPKSCES  336
Db  289  ALAFVCWRK------------------IKQSCEEENAGAEDQDG--------EGEGSKTALQPLKHSDS  331
Qy  337  SDPE  340
Db  332  KEDD  335
```

RESULT 9
US-09-651-200-4
Sequence 4, Application US/09651200
Patent No. 6429303
GENERAL INFORMATION:
APPLICANT: Green et al
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
TITLE OF INVENTION: Polypeptides Encoded Thereby
FILE REFERENCE: 15966-562 (CURA-62)
CURRENT APPLICATION NUMBER: US/09/651,200
CURRENT FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/152383
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/172909
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/183578
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
  LENGTH: 441
  TYPE: PRT
  ORGANISM: Homo sapiens
US-09-651-200-4

Query Match   5.9%;   Score 122.5;   DB 4;   Length 441;
Best Local Similarity 21.2%; Pred. No. 0.007;
Matches 77; Conservative 50; Mismatches 128; Indels 109; Gaps 17;

```
Qy   10  SGSGNEVIEGPQNATV-LKGSQARFNCTVS--QGWKL----IMWALSDMVVLSVRPMEPI  62
Db  149  SPTGAVEVQVPEDPVVALVGTDATLHCSFSPEPGFSLTQLNLIWQLTDTKQLV-------  201
Qy   63  ITNDRFTSQRYDQGGNF------------------TSEMIIHNVEPSDSGNIRCSLQNSRLHGS  108
Db  202  ---HSFTEGR-DQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFV-SIRDFGS  256
Qy  109  AYLTVQVMGELFIPSVNLV----VAENEPCEVTCLPSHWTRLP--DISWELGL---LVSH  159
Db  257  AAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPLTGN  314
Qy  160  SSYYFVPEPSDLQSAVSIILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP---QDT  216
Db  315  VTTSQMANEQGLFDVHSVLRVVLGANGTYSC------------------------------  356
Qy  217  GGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCRRRCCGCNCCC  276
Db  357  HGSVTIIGQPMTFPPEAL-----WVTVGLSVCLIALLV------------------------  389
```

```
Qy  277  RCCFCCRRKRGFRIQFQKKSEKKTNKETESGNENSGYNSDEQKTTDTASLPPKSCES  336
Db  390  ALAFVCWRK------------------IKQSCEEENAGAEDQDG--------EGEGSKTALQPLKHSDS  432
Qy  337  SDPE  340
Db  433  KEDD  436
```

RESULT 10
US-09-651-200-6
Sequence 6, Application US/09651200
Patent No. 6429303
GENERAL INFORMATION:
APPLICANT: Green et al
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
TITLE OF INVENTION: Polypeptides Encoded Thereby
FILE REFERENCE: 15966-562 (CURA-62)
CURRENT APPLICATION NUMBER: US/09/651,200
CURRENT FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/152383
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/172909
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/183578
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 6
  LENGTH: 534
  TYPE: PRT
  ORGANISM: Homo sapiens
US-09-651-200-6

Query Match   5.9%;   Score 121.5;   DB 4;   Length 534;
Best Local Similarity 21.2%; Pred. No. 0.011;
Matches 77; Conservative 49; Mismatches 129; Indels 109; Gaps 17;

```
Qy   10  SGSGNEVIEGPQNATV-LKGSQARFNCTVS--QGWKL----IMWALSDMVVLSVRPMEPI  62
Db  242  SPTGAVEVQVPEDPVVALVGTDATLRCSFSPEPGFSLAQLNLIWQLTDTKQLV-------  294
Qy   63  ITNDRFTSQRYDQGGNF------------------TSEMIIHNVEPSDSGNIRCSLQNSRLHGS  108
Db  295  ---HSFTEGR-DQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFV-SIRDFGS  349
Qy  109  AYLTVQVMGELFIPSVNLV----VAENEPCEVTCLPSHWTRLP--DISWELGL---LVSH  159
Db  350  AAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPLTGN  407
Qy  160  SSYYFVPEPSDLQSAVSIILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP---QDT  216
Db  408  VTTSQMANEQGLFDVHSVLRVVLGANGTYSC------------------------------  449
Qy  217  GGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCRRRCCGCNCCC  276
Db  450  HGSVTIIGQPMTFPPEAL-----WVTVGLSVCLIALLV------------------------  482
Qy  277  RCCFCCRRKRGFRIQFQKKSEKKTNKETESGNENSGYNSDEQKTTDTASLPPKSCES  336
Db  483  ALAFVCWRK------------------IKQSCEEENAGAEDQDG--------EGEGSKTALQPLKHSDS  525
Qy  337  SDPE  340
Db  526  KEDD  529
```

RESULT 11
US-09-651-200-24
Sequence 24, Application US/09651200
Patent No. 6429303

GENERAL INFORMATION:
APPLICANT: Green et al
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
Lymphocyte Activation Antigen B-7 Family and
TITLE OF INVENTION: Polypeptides Encoded Thereby
FILE REFERENCE: 15966-562 (CURA-62)
CURRENT APPLICATION NUMBER: US/09/651,200
CURRENT FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/152383
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/172909
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/183578
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 24
LENGTH: 534
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Sequence
OTHER INFORMATION: m25020.protein from Figure 4.
US-09-651-200-24

Query Match   Best Local Similarity   21.2%;   Score 121.5;   DB 4;   Length 534;
Matches 77;   Conservative 49;   Mismatches 129;   Indels 109;   Gaps 17;

```
Qy   10 SGSGNEVIEGPQNATV-LKGSQARFNCTVS--QGWKL----IMWALSDMVVLSVRPMEPI  62
            | : | :|: |           ||  |:  :::        ||: |: |  | ||
Db  242 SPTGAVEVQVPEDPVVALVGTDATLRCSFSPEPGFSLAQLNLIWQLTDTKQLV------  294

Qy   63 ITNDRFTSQRYDQGGNF---------------------TSEMIHNVEPSDSGNIRCSLQNSRLHGS 108
Db  295 ---HSFTEGR-DQGSAYANRTALPFDLLAQGNASLRLQRVRVADEGSFTCFV-SIRDFGS 349

Qy  109 AYILTQVMGELFIPSVNLV--VAENEPCEVTCLPSHWTRLP--DISWELGL---LVSH  159
Db  350 AAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPLIGN 407

Qy  160 SSYFVPEPSDLQSAVSIIALTPQSNGTLTCVATWKSLKARKSATVNLTVTRCP---QDT 216
Db  408 VTISQMANEQGLFDVHSVLRVVLGANGTYSC--------------------LVRNPVLQQDA 449

Qy  217 GGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCCRRRCCGCNCCC 276
Db  450 HGSVTITGQPMTFPPEAL-----WVTVGLSVCLIALLV---------------------- 482

Qy  277 RCCFCCRRKRGFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKTTDTASLPPKSCES 336
Db  483 ALAFVCWRK------------IKQSCEEENAGAEDODG------------ 525

Qy  337 SDPE 340
Db  526 KEDD 529
```

STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/633,148
FILING DATE: 16-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY ESQ., MATTHEW B.
REGISTRATION NUMBER: 39,787
REFERENCE/DOCKET NUMBER: 014618-005600US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
SEQUENCE CHARACTERISTICS:
INFORMATION FOR SEQ ID NO: 4:
LENGTH: 318 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-633-148-4

Query Match   Best Local Similarity   24.0%;   Score 118.5;   DB 2;   Length 318;
Matches 64;   Conservative 32;   Mismatches 88;   Indels 83;   Gaps 14;

```
Qy   15 EVIEGPQNATVLKGSQARFNCTVSGQ---WKLIMWALSDMVVLSVRPMEPIITNDRRTS-  70
Db  103 EIVDSASELTA--GVPNKVGTCVSEGSYPAGILSWHLDG----------KPLVPNEKGVSV 151

Qy   71 ----QRYDQGGNFT--SEMIHNVEPSDSGNIR----CSLQNSRLHGSAVLTVQVMGELF  120
Db  152 KEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRALRTAPIQPRVW 208

Qy  121 IP----SVNLVVAENEP----------------CEVTCLPSHWTRLPDISWELGLLVSHSSY 162
Db  209 EPVPLEEVQLVV----EPEGGAVAPGGTVTLTVCEVPAQPS----PQIHWMKD-------- 252

Qy  163 YFVPEPSDLQSAVSIIALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINI 222
Db  253 -GVPLPLPPSPVLILPEIGPQDQGTYSCVAHSSHPQESRAVSISIIE-PGEES------- 305

Qy  223 PGVLSSLPSLGFSLPTWGKVGLGLAGT 249
Db  306 -------------PTAGSVGGSGLGT 318
```

RESULT 12
US-08-633-148-4
Sequence 4, Application US/0863148
Patent No. 5864018
GENERAL INFORMATION:
APPLICANT: MORSER, MICHAEL J.
APPLICANT: NAGASHIMA, MARIKO
APPLICANT: HOLLANDER, DORIS A.
TITLE OF INVENTION: ANTIBODIES TO ADVANCED GLYCOSYLATION
NUMBER OF INVENTION: END-PRODUCT RECEPTOR POLYPEPTIDES AND USES THEREFOR
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP
STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
CITY: SAN FRANCISCO

RESULT 13
US-08-633-148-2
Sequence 2, Application US/0863148
Patent No. 5864018
GENERAL INFORMATION:
APPLICANT: MORSER, MICHAEL J.
APPLICANT: NAGASHIMA, MARIKO
APPLICANT: HOLLANDER, DORIS A.
TITLE OF INVENTION: ANTIBODIES TO ADVANCED GLYCOSYLATION
NUMBER OF INVENTION: END-PRODUCT RECEPTOR POLYPEPTIDES AND USES THEREFOR
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP
STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/633,148
  FILING DATE: 16-APR-1996
  CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: MURPHY ESQ., MATTHEW B.
  REGISTRATION NUMBER: 39,787
  REFERENCE/DOCKET NUMBER: 014618-00560OUS
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 326-2400
  TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 2:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 340 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: peptide
US-08-633-148-2

Query Match   5.8%;
Best Local Similarity   24.0%; Pred. No. 0.011;
Matches   64; Conservative   32; Mismatches   88; Indels   83; Gaps   14;

Qy   15 EVIEGPQNATVLKGSQARFNCTVSQG----WKLIMWALSDMVVLSVRPMEPIITNDRFTS- 70
           |||| :| | ||| | | ::| |:
Db   125 EIVDSASELTA--GVPNKWGTCVSEGSYPAGTLSWHLDG------------KPLVPNEKGVSV 173

Qy   71 ----QRYDQGGNFT--SEMIIHNVEPSDSGNIR------CSLQNSRLHGSAYLTVQVMGELF 120
                 :|| ||||||||||||||| ::| || |||||||||||
Db   174 KEQQTRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRALRTAPIQPRVW 230

Qy   121 IP----SVNLVVAENEP---------------------CEVTCLPSHWTRLPDISWELGLLVSHSSY 162
           || | | |||||||||||||||||
Db   231 EPVPLEEVLVV----EPEGGAVAPGGTVTLITCEVPAQPS----PQIHWMKD-------- 274

Qy   163 YFVPEPSDLQSAVSIILALTPQSNGTLTCVATWKSLKARKSATVNLTVRCPQDTGGGINI 222
           | |||||| : ||| : | | | | | | | ||| | || ::: | |
Db   275 -GVPLPLPPSPVLLPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIE-PGEEG---- 327

Qy   223 PGVLSSLPSLGFSLPTWGKVGLGLAGT 249
           | | | | | || | ||
Db   328 --------PTAGSVGGSGLGT 340

RESULT 14
US-09-651-200-20
Sequence 20, Application US/0965120O
Patent No. 6429303
GENERAL INFORMATION:
  APPLICANT: Green et al
  TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
  TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
  TITLE OF INVENTION: Polypeptides Encoded Thereby
  FILE REFERENCE: 15966-562 (CURA-562)
  CURRENT APPLICATION NUMBER: US/09/651,200
  CURRENT FILING DATE: 2000-08-30
  PRIOR APPLICATION NUMBER: 60/152283
  PRIOR FILING DATE: 1999-09-03
  PRIOR APPLICATION NUMBER: 60/172909
  PRIOR FILING DATE: 1999-12-21
  PRIOR APPLICATION NUMBER: 60/183578
  PRIOR FILING DATE: 2000-02-18
  NUMBER OF SEQ ID NOS: 25
  SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 20
  LENGTH: 325
  TYPE: PRT
  ORGANISM: sus sp.
US-09-651-200-20

Query Match   5.7%;
Best Local Similarity   22.9%; Pred. No. 0.014;
Matches   88; Conservative   39; Mismatches   126; Indels   132; Gaps   22;

Qy   3 LVIFLHGSSGSGNEVIEGPQNATVLKGSQARFN------------C----TVSQGWKU----IMWALSD 50
           | | |: | : || | | | ||||||
Db   10 MVLLSGAAS-------------------LK-SQAYFNETGELPCHFTNSQNLSLDELVIFWQDQD 55

Qy   51 MVVL--SVRPME-PIITNDRFTSQ-RYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLH 106
           |||| |||||| ||||||||| :||||||||||||||||||||||||||||||||
Db   56 NLVLYELYRGQEKPHVNSKYMGRISFDQA---TWTLRLHNVQIRDKGSYQCFIHHKGPH 112

Qy   107 GSAYL-----TVQVMGELFIPSVNLVV--AENEPCEVTCLPSHWTRLPDISWELGLLVSH 159
           |||| |||||||||||| : || |||| ||||||||||||||||
Db   113 GLVPIHQMSSDLSLLANFSQPEINLLTINHTENSVINLTC---------------- 151

Qy   160 SSYYFVPEPSDLQSAVSIILALTPQSNGTLTCVATWKSLKARKSATVNLTV---IRCPQDT 216
           | ||||||||||||||||||||||||||||||||||||||||||||||| ||| |
Db   152 SSTQGYPEPQ----RMYMLINTKNSTTEHDADMKKSQNNITELYNVSIRVSLPIPPET 205

Qy   217 GGGINIPGVLSSLPS--LGFSLP------------TWGKVGLGLAGTMLLTPTCT 257
           | | | : | :: | | | ||| | || |:|
Db   206 --NVSIVCVLQLEPSKTLLFSLPCNIDAKPPVQPPVPDHILW---IRAALLVTVVV 256

Qy   258 LTIRCCCCRRRCCGCNCCCRCCFCCCRRKR-------------GFRIQFQKK-SEKEKTNKET 305
           |:|||||||||||||||||||||||||| |||| || |:|:|:|||
Db   257 ----------------CGMVSEVTLRKRKRKKQPGPSNECGETIKMNRKASEQTKNRAEV 299

Qy   306 ETEGGNENSGYN-----SDEQKTTD 325
           || | | || :|||| |
Db   300 HERSDDAQCDVNILKTASDDNSTTD 324

RESULT 15
US-09-540-245A-15
Sequence 15, Application US/09540245A
Patent No. 6270984
GENERAL INFORMATION:
  APPLICANT: Goodman, Corey
  APPLICANT: Kidd, Thomas
  APPLICANT: Brose, Katja
  APPLICANT: Tessier-Lavigne, Marc
  TITLE OF INVENTION: Modulating Robo: Ligand Interactions
  FILE REFERENCE: B98-031-3
  CURRENT APPLICATION NUMBER: US/09/540,245A
  CURRENT FILING DATE: 2000-03-31
  PRIOR APPLICATION NUMBER: 60/065,544
  PRIOR FILING DATE: 1997-11-14
  PRIOR APPLICATION NUMBER: 60/081,057
  PRIOR FILING DATE: 1998-04-07
  NUMBER OF SEQ ID NOS: 20.
  SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
  LENGTH: 1395
  TYPE: PRT
  ORGANISM: Drosophila melanogaster
US-09-540-245A-15

Query Match   5.5%;
Best Local Similarity   25.8%; Pred. No. 0.2;
Matches   65; Conservative   29; Mismatches   102; Indels   56; Gaps   15;

Qy   20 PQNATVLKGSQARFNCTVSQG---WKLIMW----ALSDMVVLSVRPMEPIITNDRFTSQR 72
           |||||||||||||||||||||| |:| |:|:|||||||||||||||||||
Db   161 PKDTRVAKGETALLECGPPKGIPEPTLIWIKDGVPLDDLKAMSFGASSRV------R 211

Qy   73 YDQGGNFTSEMIIHNVEPSDSGNIRCSLQN---SRLHGSAYLTVQVMGELFIPSVNLVVA 129
           ||||||||||||||||||||||||||||| :|||||||||||||||||||||||||
Db   212 IVDGGN----LLISNVEPIDEGNYKCIAQNLVGTRESSYAKLIVQVKPYFMKEPKDQVML 267

Qy   130 ENEP----CEVTCLPSHWTRLPDISW---ELGLILVSHSSYYFVPEPSDLQSAVSIILALTP 182
           ||| | | |||||||||||| | |:|:|:|| | |||||||||||||||||

```
Db   268  YGQTATFHCSVGGDPP-----PKVLWKKEEGNIPVSRARILHD-----EKSLEISNITP  316

Qy   183  QSNGTLTCVA--TWKSLKARKSATV----NLTVIRCPQDTGGGINIPGVLSSLPSLGFSL  236
               : |||  |     |  | :  ||| | |:
Db   317  TDEGTYVCEAHNNVGQISARASLIVHAPPNFT--KRPSNKKVGLN--GVV-QLPCMASGN  371

Qy   237  PT----WGKVGL  244
              |  ||:
Db   372  PPPSVFWTKEGV  383

Search completed: April 28, 2003, 21:12:14
Job time : 18.9029 secs
```

OM protein - protein search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 16:24:31 ; Search time 38.0676 Seconds
                                            (without alignments)
                                            1337.141 Million cell updates/sec

Title:
Perfect score: 2059
Sequence: 1 MGLVIFLHGSGSGNEVIEGP............HPQASFNLASPEKVSNTTVV 382

Scoring table: BLOSUM62
                   Gapop 10.0 , Gapext 0.5
Searched: US-09-729-264-2
                   908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Post-processing: Minimum Match 0%
                 Maximum Match 100%
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
                 Listing first 45 summaries

Database :   A_Geneseq_101002:*
                1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA180.DAT:*
                2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA181.DAT:*
                3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA182.DAT:*
                4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA183.DAT:*
                5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA184.DAT:*
                6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA185.DAT:*
                7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA186.DAT:*
                8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA187.DAT:*
                9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA188.DAT:*
               10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA189.DAT:*
               11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA190.DAT:*
               12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA191.DAT:*
               13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA192.DAT:*
               14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA193.DAT:*
               15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA194.DAT:*
               16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA195.DAT:*
               17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA196.DAT:*
               18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA197.DAT:*
               19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
               20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
               21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
               21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
               22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
               23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2059	100.0	382	23	AAU75540	Human B7-like prot
2	2014	97.8	386	23	AAU75542	Human B7-like prot
3	2004	97.3	386	23	AAU75541	Human B7-like prot
4	1514	73.5	377	23	AAU75543	Human B7-like prot
5	1326	64.4	463	22	ABG28169	Novel human diagno
6	924.5	44.9	370	23	AAU75544	Mouse B7-like prot
7	583.5	28.3	631	23	AAU75547	Rat B7-like protei
8	575.5	28.2	270	23	AAU75545	Mouse B7-like prot
9	494	24.0	223	23	AAU75546	Mouse B7-like prot
10	147.5	7.2	404	22	AAB81925	Extracorporeal cir

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

RESULT 1
AAU75540
AAU75540 standard; Protein; 382 AA.

XX
ID   AAU75540 standard; Protein; 382 AA.
XX
AC   AAU75540;
XX
DT   23-APR-2002 (first entry)
XX
DE   Human B7-like protein, B7-L_h1.
XX
OS   Homo sapiens.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;

DE   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.

ALIGNMENTS

11	147.5	7.2	404	23	AAE23219	Human receptor for
12	147.5	7.2	404	23	AAU77543	Human receptor for
13	147.5	7.2	404	23	AAM48745	Human RAGE protein
14	138	6.7	592	23	ABB75751	Human immunoglobul
15	136	6.6	594	23	ABB75753	Human pancreas GP3
16	136	6.6	708	22	AAU29915	Human PRO polypept
17	136	6.6	708	23	ABB75752	Human gp354 (putat
18	135	6.6	1496	20	AAW81030	Melanoma associate
19	135	6.6	1496	21	AAY70469	Human p53 target m
20	135	6.6	1498	22	ABB11587	Human peroxidasin
21	128	6.2	1447	16	AAR68553	Human DCC protein.
22	128	6.2	1447	20	AAY33498	Human UNC-40 prote
23	128	6.2	1447	23	AAB50693	Deleted in Colorec
24	128	6.2	1728	12	AAR13144	Protein MYLK diffe
25	128	6.2	1953	23	AAU84351	Alternatively spli
26	127.5	6.2	467	19	AAW62575	Nsk2 extracellular
27	127.5	6.2	475	17	AAR94982	Alternatively spli
28	127.5	6.2	863	19	AAW62569	Mouse receptor tyr
29	127.5	6.2	867	19	AAR62583	Nsk2 receptor. Mu
30	127.5	6.2	871	17	AAR84087	Mouse receptor tyr
31	127.5	6.2	871	19	AAW62568	Nsk2 receptor with
32	127.5	6.2	873	17	AAR84092	Alternatively spli
33	127.5	6.2	873	19	AAW62573	Nsk2 receptor with
34	127.5	6.2	881	17	AAR84091	Nsk2 receptor with
35	127.5	6.2	881	19	AAW62572	Nsk2 (altern
36	126.5	6.1	869	18	AAW26611	Mouse muscle-speci
37	126.5	6.1	869	18	AAW26506	Human Dmk receptor
38	126.5	6.1	869	22	AAG77857	Mutant protein mMu
39	125.5	6.1	537	22	AAG77858	Mutant protein mMu
40	125.5	6.1	576	22	AAG77856	Protein of muscle
41	125.5	6.1	869	22	AAB68421	Amino acid sequenc
42	124	6.0	1070	18	AAW08747	Human colon carcin
43	124	6.0	1070	22	ABG61942	prostate cancer-as
44	124	6.0	1070	23	AAO17355	Human transmembran
45	123.5	6.0	344	22	AAM38713	Human polypeptide

Left column

```
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13028.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 1; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I)
CC   and polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation that
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of human B7-L_h1.
XX
SQ   Sequence   382 AA;

     Query Match             100.0%;  Score 2059;  DB 23;  Length 382;
     Best Local Similarity   100.0%;
     Matches  382;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1  MGLVIFLHGSGSGNEVIEGPQNATVLKGSQOARNCTVSQGWKLIMWALSDMVVLSVRPME   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1  MGLVIFLHGSGSGNEVIEGPQNATVLKGSQOARNCTVSQGWKLIMWALSDMVVLSVRPME   60

Qy      61  PIITNDRFTSQRYDQGGNETSEMIIHNYEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELF  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61  PIINDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAVLTVQVMGELF   120

Qy     121  IPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILAL  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121  IPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILAL  180

Qy     181  TPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWG  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181  TPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWG  240

Qy     241  KVGLGLAGTMLLTPTCLTIRCCCCRRRCCGCNCCCCRCCPCCRRKRGFRIQFQKKSEKEK  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241  KVGLGLAGTMLLTPTCLTIRCCCCRRRCCGCNCCCCRCCPCCRRKRGFRIQFQKKSEKEK  300
```

Right column

```
RESULT 2
AAU75542
ID   AAU75542 standard; Protein; 386 AA.
XX
AC   AAU75542;
XX
DT   23-APR-2002 (first entry)
XX
DE   Human B7-like protein, B7-L_h3.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Homo sapiens.
XX
XX
PN   WO2020070710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13030.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 3; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I).
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as

Qy     301  TNKETETESGNENGGYNSDEQKITDTASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRP  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301  TNKETETESGNENGGYNSDEQKITDTASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRP  360

Qy     361  ASHPQASFNLASPEKVSNTTVV  382
            ||||||||||||||||||||||
Db     361  ASHPQASFNLASPEKVSNTTVV  382
```

Left column

```
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid) endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of human B7-L_h3.
XX
SQ   Sequence   386 AA;

Query Match            99.7%;  Score 2014;  DB 23;  Length 386;
Best Local Similarity  99.7%;  Pred. No. 2.9e-158;
Matches  373;  Conservative   1;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    9 GSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 68
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13 GSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72

Qy   69 TSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLIVQVMGELFIPSVNLVV 128
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 TSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLIVQVMGELFIPSVNLVV 132

Qy  129 AENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDSDLQSAVSILALTPQSNGTL 188
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133 AENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDSDLQSAVSILALTPQSNGTL 192

Qy  189 TCVATWKSLKARKSATVNLTVIRCPQDTGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG 248
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193 TCVATWKSLKARKSATVNLTVIRCPQDTGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG 252

Qy  249 TMLLITPTCLTLIRCCCCRRRCCGNCNCCCRCCFCCRRKRGFRIQFQKKSEKEKTNKETETE 308
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  253 TMLLITPTCLTLIRCCCCRRRCCGNCNCCCRCCFCCRRKRGFRIQFQKKSEKEKTNKETETE 312

Qy  309 SGNENSGYNSDEQKITTDTASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASF 368
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  313 SGNENSGYNSDEQKITTDTASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASF 372

Qy  369 NLASPEKVSNTTVV 382
         |||||||||||||
Db  373 NLASPEKVSNTTVV 386


RESULT 3
AAU75541
ID   AAU75541 standard; Protein; 386 AA.
XX
AC   AAU75541;
XX
DT   23-APR-2002 (first entry)
XX
DE   Human B7-like protein, B7-L_h2.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Homo sapiens.
XX
PN   WO200200710-A2.
```

Right column

```
XX
03-JAN-2002.
XX
PD   28-JUN-2001; 2001WO-US20719.
XX
PF   28-JUN-2000; 2000US-214512P.
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA,  Sarmiento UM,  Schultz HJ,  Chute HT;
XX
WPI; 2002-130881/17.
N-PSDB; ABK13029.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 2; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising the
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle hyperplasia in transgenic mice
CC   overexpressing B7-L lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of human B7-L_h2.
XX
SQ   Sequence   386 AA;

Query Match            99.5%;  Score 2004;  DB 23;  Length 386;
Best Local Similarity  97.3%;  Pred. No. 2e-157;
Matches  372;  Conservative   1;  Mismatches   2;  Indels   0;  Gaps   0;

Qy    9 GSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 68
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   13 GSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72

Qy   69 TSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLIVQVMGELFIPSVNLVV 128
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 TSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLIVQVMGELFIPSVNLVV 132

Qy  129 AENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDSDLQSAVSILALTPQSNGTL 188
```

```
Db  133  AENPCEVTCLPSHWTWLPDISWELGLLIVSHSSYYFVPEPSDLQSAVSILALTPQSNGTL
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Qy  189  TCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG  248
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193  TCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG  252
Qy  249  TMLLTPTCTLTIRCCCCRRCCGCNCCCRRCCFCCRRKRGFRIQFQKKSEKEKTNKETETE  308
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db  253  TMLLTPTCTLTIRCCCCRRCCGCNCCCRRCCFCCRRKRGFRIQFQKKSEKEKTNKETETE  312
Qy  309  SGNENSGYNSDEQKTTDTASLPPPKSCESSDPEQRNSSCGPPHQRADQRPPRASHPQASF  368
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  313  SGNENSGYYNSDEQKTTDTASLPPPKSCESSDPEQRNSSCGPPHQRADQRPPRASHPQASF  372
Qy  369  NLASPEKVSNTTVV  382
              |||||||||||||
Db  373  NLASPEKVSNTTVV  386
```

RESULT 4

AAU75543

ID   AAU75543 standard; Protein; 377 AA.
AC   AAU75543;
XX
DT   23-APR-2002 (first entry)
XX
DE   Human B7-like protein, B7-L_h4.
XX
OS   Homo sapiens.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Homo sapiens.
XX

PN   WO200200710-A2.

PD   03-JAN-2002.

PF   28-JUN-2001; 2001WO-US020719.
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-072926A.

PA   (AMGE-) AMGEN INC.

PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;

DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13031.

PT   New B7-like polypeptides and polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PS   Claim 13; Fig 4; 135pp; English.

Claim 13; Fig 4; 135pp; English.

The invention relates to an isolated B7-like (B7-L) polypeptide (I).
The polypeptide, polynucleotide encoding it and antibody against (I) are
useful for treating B7-like polypeptide-related disease, disorders or
conditions including reproductive disorders (e.g. infertility,
miscarriage, preterm labour and delivery and endometriosis or
proliferative disorders. Antibodies, soluble proteins comprising
extracellular domains and other regulators of B7-L polypeptides are
useful for enhancing the immune response to tumours. (I) plays a role in
growth and maintenance of cancer cells based on the observation of
seminal vesicle hyperplasia in transgenic mice overexpressing B7-L.

CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of human B7-L_h4.
XX
SQ   Sequence   377 AA;

Query Match   73.5%;   Score 1514;   DB 23;   Length 377;
Best Local Similarity   88.2%;   Pred. No. 6.5e-117;
Matches 284;   Conservative 12;   Mismatches 20;   Indels 6;   Gaps 2;

```
Qy    9  GSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF   68
             |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   13  GSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF   72
Qy   69  TSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV  128
             ||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db   73  TSQRYDQGGNLTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV  132
Qy  129  AENPCEVTCLPSHWTRLPDISWELGLLIVSHSSYYFVPEPSDLQSAVSILALTPQSNGTL  188
             ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  133  AENPCEVTCLPSHWTWLPDISWELGLLIVSHSSYYFVPEPSDLQSAVSILALTPQSNGTL  192
Qy  189  TCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG  248
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193  TCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG  252
Qy  249  TMLLTPTCTLTIRCCCCRRCCGCNCCCRRCCFCCRRKRGFRIQFQKKSEKEKTNKETETE  308
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  253  TMLLTPTCTLTIRCCCCRRCCGCNCCCRRCCFCCRRKRGFRIQFQKKSEKEKTNKETETE  308
Qy  309  SGNENSGYNSDEQKTTDTASLP  330
             ::|: ::||:|||
Db  310  T--PATIQMNKRPQTPLLSLP  328
```

RESULT 5

ABG28169

ID   ABG28169 standard; Protein; 463 AA.
AC   ABG28169;
XX
DT   18-FEB-2002 (first entry)
XX
DE   Novel human diagnostic protein #28160.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX

RESULT 6

Left column:

PN   WO20017S067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS92356.
XX
PS   Claim 20; SEQ ID No 58528; 103pp; English.
XX

The invention relates to isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity. (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences are useful in medical diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders and products dependent on DNA and amino acid sequences. ABG00010-ABG30377 represent novel human amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ   Sequence   463 AA;

Query Match   64.4%; Score 1326; DB 22; Length 463;
Best Local Similarity   100.0%; Pred. No. 3e-101;
Matches 243; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy   45 MWALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSR 104
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MWALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSR 60

Qy  105 LHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYF 164
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYF 120

Qy  165 VPEPSDLQSAVSILALTPQSNGTLTICVAFWKSLKARKSATVNLTVIRCPQDTGGGINIPG 224
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 VPEPSDLQSAVSILALTPQSNGTLTICVAFWKSLKARKSATVNLTVIRCPQDTGGGINIPG 180

Qy  225 VLSSLPFSLGFSLPTWGKVGLGLAGTMLLTPPCTLTIRCCCCRRRCGCNCCCRCCFCCRR 284
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 VLSSLPFSLGFSLPTWGKVGLGLAGTMLLTPPCTLTIRCCCCRRRCGCNCCCRCCFCCRR 240

Qy  285 KRG 287
         |||
Db  241 KRG 243
```

Right column:

AAU75544
ID   AAU75544 standard; protein; 370 AA.
XX
AC   AAU75544;
XX
DT   23-APR-2002 (first entry)
XX
DE   Mouse B7-like protein, B7-L_m1.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological; antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic; antiinflammatory; dermatological; antipsoriatic; neuroprotective; antiasthmatic; haemostatic; antithyroid; antiulcer; antiallergic; antidiabetic; nephrotropic; antibacterial; virucide; tumour; cancer; reproductive disorder; graft versus host disease; autoimmune disease; toxic shock syndrome; allergy; nephropathy; skin disorder; endocrinopathy; lymphoproliferative disorder.
XX
OS   Mus musculus.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13032.
XX
PS   Claim 13; Fig 5; 135pp; English.

New B7-like polypeptides, polynucleotides and their modulators, useful for diagnosing, preventing and treating reproductive, immune and proliferative disorders, e.g. cancer and arteriosclerosis.

The invention relates to an isolated B7-like (B7-L) polypeptide (I). The polypeptide, polynucleotide encoding it and antibody against (I) are useful for treating B7-like polypeptide-related disease, disorders or conditions including reproductive disorders (e.g. infertility), cancer including seminal vesicle cancer, lung, brain, breast, ovarian, testicular cancer and cancers of haematopoietic system. B7-L polypeptide pathway can be manipulated to regulate cytotoxic T-lymphocyte response in allograft transplantation, graft versus host disease. T-cell dependent B-cell mediated diseases and autoimmune diseases. B7-L molecules are useful for alleviating the symptoms associated with diseases involving chronic immune cell dysfunction or to treat autoimmune diseases such as systemic lupus erythematosus, rheumatoid arthritis, multiple sclerosis, diabetes, immune thrombocytopenic purpura and psoriasis, chronic inflammatory disease such as inflammatory bowel disease (Crohn's disease and ulcerative colitis), Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They are also useful as immunosuppressive agents for bone marrow and organ transplantation or to prolong graft survival. B7-L molecules are also useful for diagnosis and treatment of diseases involving abnormal cell proliferation, including arteriosclerosis and vascular restenosis. Antagonists of B7-L polypeptides are useful for alleviation of toxic shock syndrome or allosensitisation due to blood transfusions, and for treatment of allergy, asthma and hypersensitivity reactions, skin disorders (e.g. glomerulonephritis), skin disorders (penphigus and nephropathies (e.g. glomerulonephritis).

The invention also relates to soluble proteins comprising extracellular domains and other regulators of B7-L polypeptides are useful for enhancing the immune response to tumours. (I) plays a role in growth and maintenance of cancer cells based on the observation of seminal vesicle hyperplasia in transgenic mice overexpressing B7-L polypeptide. Hence modulators of (I) are useful for the treatment of cancer.

```
CC    pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC    (extrinsic alveolitis) vasculopathies, coeliac disease, anaemia,
CC    thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC    lymphoproliferative disorders such as multiple myeloma. The present
CC    sequence represents the amino acid sequence of mouse B7-L_m1.
XX
SQ    Sequence  370 AA;

Qy    Query Match    44.98;  Score 924.5; DB 23; Length 370;
Best Local Similarity  50.8%;  Pred. No. 3.6e-68;
Matches 197;  Conservative  50;  Mismatches 104;  Indels  37;  Gaps  6;

Qy       3 LVIF--LHGSGSGNEVIEGPQONATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPME 60
           ||||  :||||||||||:||||||:||||||:|||||:|||||||:|:||.::|||||||||
Db      12 LVILAQLTASGSSYQIIEGPQNVTVLKDSEAHFNCTVIHGWKLLMWTLNQMVVLSLTTQG 71

Qy      61 PIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSDSGNIRCSLQNSRLHGSAYLIYQVMGELF 120
           ||:|||:||||||:||||||||||||||:|||||||||||:||||||||||:||:|||||||
Db      72 PIITNNRFTYASYNSTDSFISELIIHDVQPSDSGSVQCSLQNSHGFGSAFLSVQVMGTLN 131

Qy     121 IPSVNLIVVAAENEPCEVTCLPSHWTRLPDISWELGLLIVSHSSYYFVPEPSDLQSAVSIIAL 180
           |||:||:|||||||:|||||||:||||:||||||||||:|||:|||:||||||:.||||||||
Db     132 IPSNNLIVTEGEPCNVTCYAVGWTSLPDISWELEVPVSHSSTNSFLEPGNFMRVLSVLDL 191

Qy     181 TPQSNGTLTCVAIFWKSLRARKSATVNLIVIRCPQDTGGININPGVLSSLPSLGFSLPTWG 240
           :|||||||||:||||||||||||||||:||||:|||||||||||||||:||||||||||||
Db     192 TPLGNGTLTCVAELKDLQASKSLTVNLTVVQPPPD----------------SIGEEGPALPTWA 239

Qy     241 KVGLGLLAGTMLLTPTCTLTIRCCCCRRRCCGCNCCCCFCCCRRKRGFRIQFQ--KKSE 297
           |:||:|||||||||||||||:|.||||  |||
Db     240 IILLAVAFSILLLILIIVLIIIFCCC------------------CASRREKEESTYQNEIRKSA 284

Qy     298 KEKTNK---ETFTESGNENSGYNSDEQKTTDTASLPPKSCESSDPEQRNSSCGPPHQRAD 354
           :|||||   |||
Db     285 NMRTNKADPEFKLKSGKENYGSDDEAKAAQTASLPPKSAEVSLPPEKRSSSL--PYQELN 342

Qy     355 QRPPRPASHPQASFNLASPEKVSNTTVV 382
           |:|||||||:||:||:|||||||:||||
Db     343 KHQPGPATHPRVSFDIASPQKVRNVTLV 370
```

```
RESULT 7
AAU75547

ID    AAU75547 standard; protein; 631 AA.
XX
AC    AAU75547;
XX
DT    23-APR-2002  (first entry)
XX
DE    Rat B7-like protein, B7-1.
XX
KW    Rat; B7-like protein; B7-1; antiinfertility; gynaecological;
KW    antitumour; cytostatic; immunosuppressive; antiarthritic;
KW    antiinflammatory; dermatological; antipsoriatic; antirheumatic;
KW    antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW    antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW    reproductive disorder; graft versus host disease; autoimmune disease;
KW    toxic shock syndrome; allergy; nephropathy; skin disorder;
OS    endocrinopathy; lymphoproliferative disorder.
OS    Ratus rattus.
XX
PN    WO200200710-A2.
XX
PD    03-JAN-2002.
XX
PF    28-JUN-2001; 2001WO-US20719.
PR    28-JUN-2000; 2000US-214512P.
PR    28-NOV-2000; 2000US-0729264.
XX
PA    (AMGE-) AMGEN INC.
XX
```

```
PI    Welcher AA,  Sarmiento UM,  Schultz HJ,  Chute HT;
XX
DR    WPI; 2002-130881/17.
XX
PT    New B7-like polypeptides, polynucleotides and their modulators, useful
PT    for diagnosing, preventing and treating reproductive, immune and
PT    proliferative disorders, e.g. cancer and arteriosclerosis -
PS    Disclosure; Fig 8; 135pp; English.
XX
CC    The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC    The polypeptide, polynucleotide encoding it and antibody against (I)
CC    useful for treating B7-like polypeptide-related disease, disorders or
CC    conditions including reproductive disorders (e.g. infertility,
CC    miscarriage, preterm labour and delivery and endometriosis) and
CC    proliferative disorders. Antibodies, soluble proteins comprising
CC    extracellular domains and other regulators of B7-L polypeptides are
CC    useful for enhancing the immune response to tumours. (I) plays a role in
CC    growth and maintenance of cancer cells based on the observation of
CC    seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC    polypeptide. Hence modulators of (I) are useful for the treatment of
CC    cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC    testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC    pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC    in allograft transplantation, graft versus host disease, T-cell
CC    dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC    molecules are useful for alleviating the symptoms associated with
CC    diseases involving chronic immune cell dysfunction or to treat
CC    autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC    arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC    purpura and psoriasis, chronic inflammatory disease such as
CC    inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC    Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC    are also useful as immunosuppressive agents for bone marrow and organ
CC    transplantation or to prolong graft survival. B7-L molecules are also
CC    useful for diagnosis and treatment of diseases involving abnormal cell
CC    proliferation, including arteriosclerosis and vascular restenosis.
CC    Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC    shock syndrome or allosensitisation due to blood transfusions, and for
CC    treatment of allergy, asthma and hypersensitivity reactions,
CC    nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC    pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC    (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC    thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC    lymphoproliferative disorders such as multiple myeloma. The present
CC    sequence represents the amino acid sequence of rat B7-1.
XX
SQ    Sequence  631 AA;

Qy    Query Match    28.38;  Score 583.5; DB 23; Length 631;
Best Local Similarity  26.98;  Pred. No. 1.2e-39;
Matches 173;  Conservative  43;  Mismatches 111;  Indels  317;  Gaps  11;

Qy       3 LVIF--LHGSGSGNEVIEGPQ 22
           |||   ||| ||:|||||
Db      12 LVILAQLTASGSSYQIIEGPQMAYSCQPLQESPLLGFPRLRFIHLFVLLLVGLLQISSGI 71

Qy      22 --------NATIVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF---- 68
                   |:|| |:|||.|:|||| ||||||:||| |:|||||:||||||||||
Db      72 VGQVSKSVRNVTVLKDSEAHFNCTVTHGWKLLMWTLNQMVVLSLTTQGPIITNNRFEKAL 131

Qy      69 ----------------------TSQRYDQGGNFTSEMI 84
                                 : ||:||| ||||||
Db     132 LSCDYKFCSEEQSIHRIWQKHDKMVLSVISGVPEWPKYXKNRTTVASYNSTDSFISELI 191

Qy      85 IHNVEPSDSDSGNIRCSLQNSRLHGSAYL------------ 112
           |:|:||||| ::||||||||||||||||
Db     192 IHDVQPSDSGSVQCSLQNSHGFGSAFLSVQVYDIANNYSFSLLGLILSDRGTYTCVQVQRY 251

Qy     112 ----------TVQVMGELFIPSVNLVVAAENEPCEVTCLPSHWTRLPDISWELGLLVSHS 160
                   :||:||||| || |:|| |:|||||:|||||||:||:||:|||||||||||
Db     252 EGGSYVVKHLTTVEVWGTLNIPSNNLIVTEGEPCNVTCYAVGWTSLPDISWELEVPVSHS 311
```

```
Qy  161 ------------------------------------------------SIYFVPEPSDLQ 172
Db  312 LSVRADFPPTPNITEYGNPSADIKRITCFASGGFPKPRLSWLENGRELNSYNSFLEPGNFM 371

Qy  173 SAVSILALTPQSNGTLTCVATWKSLKARKSATVNL-------------------------- 207
        :|:|||| |||| ||||||||||||||||||
Db  372 RVLSVLDIFPLGNGTLTCVAELKDLQASKLIVNLGINTTISQDPESELYITSSQLDFNA 431

Qy  208 --------------TVIRCPQDTGGGINIPGVLSLLPSLGFSLPTWGKVGLGL 246
                      ||:|| |||| ::| | ||| | |
Db  432 TYDHFTDCFIEYGDAHVSQNFTVVQPPPD---------SIGEEGPALPTWAIILLAV 479

Qy  247 AGTMLL-------------------------------------------TPT 255
        |::||
Db  480 AFSLLILIIVLIIIFTWYRPPEDPDDEKQTVPEPAWAGPDAVKAIIFFIAITVIAVIAA 539

Qy  256 CTLTIRCCCRRRCCGCNCCCRCCFCCRRKRGFRIQFQ-KKSEKEKTNK---ETETESGN 311
        |: ||| | |  || |||  ||:|| :|| ||||  ||:
Db  540 IAIIFCCCASR------------REKEESTYQNEIRKSANMRTNKADPETKLKSGK 585

Qy  312 ENSGVNSDEQKTTDTASLPPKSCESSSDPEQRNSSCGPPHQRADQ 355
        ||:|||||| |||:|||:|| |||
Db  586 ENYGYSSDECITVKFRRCFRRRNEASRETNKNLYIGPVEEAAAEQ 629
```

RESULT 8
AAU75545

ID   AAU75545 standard; Protein; 270 AA.
XX
AC   AAU75545;
XX
DT   23-APR-2002 (first entry)
XX
DE   Mouse B7-like protein, B7-L_m2.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Mus musculus.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13033.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 6; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusion, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of mouse B7-L_m2.
XX
SQ   Sequence 270 AA;

Query Match   28.0%; Score 575.5; DB 23; Length 270;
Best Local Similarity   35.1%; Pred. No. 1.8e-39;
Matches 136; Conservative 42; Mismatches 73; Indels 137; Gaps   6;

```
Qy    3 LVIF--LHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPME 60
        ||| || :||||||||| |||| |:|||||||:||
Db   12 LVILAQLIRASGSSYQIIEGPQNTVLKDSEAHFNCTVTHGWKLLMWLTLNQMVVLSLTTQG 71

Qy   61 PIITNDRFTSQRYDQGGNFTSEMTIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVWGELF 120
        ||||||:|: |: |||||| ||::::||||
Db   72 PIITNNRFTTASYNSTDSFISELIIHDVQPSDSGSVQCSIQNSHGFGSAFLSVQ----- 125

Qy  121 IPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSTLAL 180
Db  126 ------------------------------------------------------------ 125

Qy  181 TPQSNGTLTCVATWKSLKARKSATVNLIVIRCPQDTGGGINIPGVLSLLPSLGFSLPTWG 240
        |||||||||||||||||||||||  |:||||||||||||||||||  ::||
Db  126 ------------------------------------------DSIGEEGPALPTWA 139

Qy  241 KVGLGLAGTMLLTPTCTLTIRCCCCRRRCCGCNCCCRCCFCCRRKRGFRIQFQ--KKSE 297
         :|: |||||||||||||||||||||||||||||||||||  ||||
Db  140 IILLAVAFSLILILIIVLIIIFCCC--------------------CASREKEESTYQNEIRKSA 184

Qy  298 KEKTNK---ETETESGNENSGYNSDEQKTTDTASLPPKSCESSSDPEQRNSSCGPPHQRAD 354
         |||||  |||||||||||||||||||||||||||||||||||||||||||||||
Db  185 NMRTNKADPETKLKSGKENYGYSSDEAKAAQTASLPPKSAEVSLPEKRSSSSL--PYQELN 242

Qy  355 QRPPRPASHPQASFNLASPEKVSNTTVV 382
        : |||: :||:|:|:|||:|||:: |:|
Db  243 KHOPGPATHPRVSFDIASPQKVRNVTLV 270
```

RESULT 9
AAU75546
ID   AAU75546 standard; Protein; 223 AA.
XX
AC   AAU75546;
XX

DT   23-APR-2002 (first entry)
XX
DE   Mouse B7-like protein, B7-L_m3.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Mus musculus.
XX
PN   WO2002007 10-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130801/17.
DR   N-PSDB; ABK13034.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 7; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising and
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of mouse B7-L_m3.

XX
SQ   Sequence   223 AA;
XX

```
Query Match         24.0%;  Score 494;  DB 23;  Length 223;
Best Local Similarity   31.9%;  Pred. No. 7.7e-33;
Matches 123;  Conservative 37;  Mismatches 47;  Indels 178;  Gaps 7;

Qy      3 LVIF--LHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWRKLIMWALSDNVVLSVRPME  60
            ||||  ::|||   |||||||||:||||:|||||||:||| ||| |: |||:|:||||:|
Db     12 LVILAQLTASGSSYQIIEGPQNVTVLKDSEAHFNCTVTHGWKLLAMWTLNQMVVLSITTQG  71

Qy     61 PIITNDRTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVNGELF   120
            ||||| || :  | :|||| ||||:| |||||||:: |||||: |||||:|:|:|
Db     72 PIITNNRFTYASYNSTDSFISELIIHDVQPSDSGSSVQCSLQNSHGFGSAFLSVQ------  125

Qy    121 IPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILAL   180
                                             :|:  :    ESTY-----QNEI-----
Db    126 ------------------------------------ESTY-----QNEI-----       133

Qy    181 TPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWG   240
               :||   ||
Db    134 ----------RKSANM-                                            139

Qy    241 KVGLGLLAGTMLLTPYCTLTIRCCCCRRRCCGCNCCCRCCPCCRRKRGFRIQFQKKSEKEK  300
                                                       :           |     R
Db    140 ------------------------------------------               R    140

Qy    301 TNK---ETETESGNENSGINSDEQKTTDFIASLPPKSCESSDPEQRNSSCGPPHQRADQRP  357
            |||   |: |       :  :|| ||| ||||||||  ||  |:  ::       |  :
Db    141 TNKADPETKILKSGKENYGYSSDEAKAAQTASLPPKSAEVSLPEKRSSSL--PYQELNKHQ  198

Qy    358 PRPASHPQASFNLASPEKVSNTTVV                                     382
            | |:||| |||::|||:||.|:|:|
Db    199 PGPATHPRVSFDIASPQKVRNVTLV                                     223
```

RESULT 10
AAB81925
ID   AAB81925 standard; protein; 404 AA.
XX
AC   AAB81925;
XX
DT   15-JUN-2001 (first entry)
XX
DE   Extracorporeal circulation material receptor protein.
XX
KW   Extracorporeal circulation; carbonyl stress product; receptor;
KW   diabetes; vascular lesion; excretory dysfunction.
XX
OS   Unidentified.
XX
PN   WO200118060-A1.
XX
PD   15-MAR-2001.
XX
PR   08-SEP-2000; 2000WO-JP06172.
PR   08-SEP-1999; 99JP-0254463.
XX
PA   (TORA) TORAY IND INC.
XX
PI   Shimizu S, Kubota M, Akiyama H, Usui M;
XX
DR   WPI; 2001-290314/30.
XX
PT   Material for extracorporeal circulation, applicable in selective
PT   elimination of diabetic complication factors such as carbonyl stress
PT   products caused by abnormally promoted carbonyl stress from excretory
PT   dysfunction in vascular lesions -
XX
PS   Claim 1; Page 31-32; 36pp; Japanese.
XX
CC   The present invention describes a material for extracorporeal circulation

CC which is made from a water-insoluble carrier immobilized with a protein
CC having the sequence shown here. The materials of the invention, including
CC adsorbents, are for extracorporeal circulation, which are applicable in
CC the selective elimination of diabetic complication factors from a body
CC fluid, and are therefore useful in treating vascular lesions like
CC arteriosclerosis due to carbonyl stress products caused by abnormally
CC promoted carbonyl stress from excretory dysfunction.
XX
SQ Sequence   404 AA;

Query Match   7.2%;   Score 147.5;   DB 22;   Length 404;
Best Local Similarity   22.6%; Pred. No. 0.00076;
Matches   85; Conservative   40; Mismatches 114; Indels 137; Gaps   18;

Qy   15 EVIEGPQNATVLKGSQARFNCTVSQG---WKLIMWALSDMVVLSVRPMEPIITNDRFTS- 70
             |:|| |:|:: :| |:|: | ::|: |
Db   125 EIVDSASELTA--GVPNKVGTCVSEGSYPAGTLSWHLDG-------KPLVPNEKGVSV 173

Qy   71 ----QRYDQGGNFT--SEMIHNVEPSDSGNIR-----CSLQNSRLHGSAYLTVQVMGELF 120
             :| : | | |:: |:|
Db   174 KEQTRHPETGLFTLQSELM---VTPARGGDPRFTFSCSFSPGLPRHRALRTAPIQPRVW 230

Qy   121 IP-----SVNLVVAENEP----------------------CEVTCLPSHWTRLPDISWELGLLVSHSSY 162
            | | | | ||| |||
Db   231 EPVPLEEVQLVV---EPEGGAVAPGGTVTLTCEVPAQPS-----PQIHWMKD------- 274

Qy   163 YFVPEPSDSLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINI 222
             | || | | | | ||||||| :: | : |
Db   275 -GVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIE-PGEEG---- 327

Qy   223 PGVLSSLPSIGFSLPTWGKVGLGLAGTMLLT------------PTCTLTIRCCCCRRRCCGCNCC 275
             | | || | | | || :|
Db   328 ------PTAGSVGGSGLGTLALALGILGGLGTAALLIGVILWQRR------- 366

Qy   276 CRCCFCCRRKRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQKTTDTASLPPKS 333
             ::|: |: : | | | |:| ||
Db   367 --------QRRGEERKAPENQEEEERAELN----------------- 389

Qy   334 CESSDPEQRNSSCGPP 349
             ||||| | | | |
Db   390 -QSEEPEAGESSTGGP 404

RESULT 11
AAE23219
ID   AAE23219 standard; Protein; 404 AA.
XX
AC   AAE23219;
XX
DT   27-AUG-2002 (first entry)
XX
DE   Human receptor for advanced glycosylation end product (RAGE) protein.
XX
KW   Human; Receptor for advanced glycosylation end product; RAGE; cardiant;
KW   tissue growth; neointimal formation; blood vessel; restenosis; diabetes;
KW   myocardial infarction; angioplasty; peripheral vascular surgery; angina;
KW   transgenic animal; acute thrombotic stroke; venous thrombosis.
XX
OS   Homo sapiens.
XX
PN   WO20020889-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US320036.
XX
PR   13-OCT-2000; 2000US-0687528.
XX
PA   (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI   Stern DM, Schmidt A, Marso S, Topol E, Lincoff AM;
XX
DR   WPI: 2002-426260/45.

DR   N-PSDB; AAD36952.
XX
PS   Disclosure; Page 16; 43pp; English.
PT   product.-
PT   comprises administering inhibitor of receptor for advance glycation end
PT   of subject suffering from diabetes, stroke and preventing restenosis,
PT   Inhibiting new tissue growth or neointimal formation in blood vessels
XX

SQ Sequence   404 AA;

The invention relates to a method for inhibiting new tissue growth or
neointimal formation in blood vessels in a subject that has experienced
blood vessel injury and preventing exaggerated restenosis in a diabetic
subject. The method comprises administering an inhibitor of receptor for
advanced glycation/glycosylation end product (RAGE), so as to inhibit new
tissue growth or neointimal formation in subject's blood vessels and
preventing restenosis in the subject. The method is useful for inhibiting
new tissue growth or neointimal formation in blood vessels in a subject
like non-human animal, a transgenic non-human animal or a human suffering
from diabetes, acute thrombotic stroke, venous thrombosis, unstable
angina, myocardial infarction, abrupt closure following angioplasty or
stent placement, or thrombosis as a result of peripheral vascular surgery
The method is also useful for preventing restenosis and for determining
whether a compound inhibits new tissue growth in a blood vessel in a
subject. The present sequence is human receptor for advanced
glycosylation end product (RAGE) protein.

Query Match   7.2%;   Score 147.5;   DB 23;   Length 404;
Best Local Similarity   22.6%; Pred. No. 0.00076;
Matches   85; Conservative   40; Mismatches 114; Indels 137; Gaps   18;

Qy   15 EVIEGPQNATVLKGSQARFNCTVSQG---WKLIMWALSDMVVLSVRPMEPIITNDRFTS- 70
             |:|| |:|:: :| |:|: | ::|: |
Db   125 EIVDSASELTA--GVPNKVGTCVSEGSYPAGTLSWHLDG-------KPLVPNEKGVSV 173

Qy   71 ----QRYDQGGNFT--SEMIHNVEPSDSGNIR-----CSLQNSRLHGSAYLTVQVMGELF 120
             :| : | | |:: |:|
Db   174 KEQTRHPETGLFTLQSELM---VTPARGGDPRFTFSCSFSPGLPRHRALRTAPIQPRVW 230

Qy   121 IP-----SVNLVVAENEP----------------------CEVTCLPSHWTRLPDISWELGLLVSHSSY 162
            | | | | ||| |||
Db   231 EPVPLEEVQLVV---EPEGGAVAPGGTVTLTCEVPAQPS-----PQIHWMKD------- 274

Qy   163 YFVPEPSDSLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINI 222
             | || | | | | ||||||| :: | : |
Db   275 -GVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIE-PGEEG---- 327

Qy   223 PGVLSSLPSIGFSLPTWGKVGLGLAGTMLLT------------PTCTLTIRCCCCRRRCCGCNCC 275
             | | || | | | || :|
Db   328 ------PTAGSVGGSGLGTLALALGILGGLGTAALLIGVILWQRR------- 366

Qy   276 CRCCFCCRRKRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQKTTDTASLPPKS 333
             ::|: |: : | | | |:| ||
Db   367 --------QRRGEERKAPENQEEEERAELN----------------- 389

Qy   334 CESSDPEQRNSSCGPP 349
             ||||| | | | |
Db   390 -QSEEPEAGESSTGGP 404

RESULT 12
AAU77543
ID   AAU77543 standard; Protein; 404 AA.
XX
AC   AAU77543;
XX
DT   05-JUN-2002 (first entry)
XX
DE   Human receptor for advanced glycation end product (RAGE).
XX
KW   Receptor for advanced glycation end product; RAGE; receptor;
KW   amyloid beta peptide; blood-brain barrier; neurovascular stress;

KW   cerebral vasoconstriction suppressor; cerebral blood flow enhancer;
KW   cerebral amyloid angiopathy; transgenic animal; amyloid beta precursor;
KW   Alzheimer's disease; Down's syndrome; head trauma; stroke; human.
XX
OS   Homo sapiens.
XX
PN   WO20021Á519-A1.
XX
PD   21-FEB-2002.
XX
PF   14-AUG-2001; 2001WO-US25416.
XX
PR   14-AUG-2000; 2000US-0638648.
XX
PA   (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI   Stern DM, Schmidt AM, Yan SD, Zlokovic B;
XX
DR   WPI; 2002-257610/30.
DR   N-PSDB; ABK10856.
XX
PS   Disclosure; Page 16; 68pp; English.
XX
CC   The invention describes a method of ameliorating neurovascular stress,
CC   and decreasing cerebral vasoconstriction in subject suffering from
CC   chronic or acute cerebral amyloid angiopathy, comprising administering
CC   an inhibitor (I) of receptor for advanced glycation end product (RAGE).
CC   (I) inhibits transcytosis of amyloid beta peptides across blood-brain
CC   barrier, thus decreasing cerebral vasoconstriction and increasing
CC   cerebral blood flow. (I) is useful for treating amyloid angiopathy in a
CC   subject, decreasing cerebral vasoconstriction in a transgenic non-human
CC   animal (preferably, transgenic mouse overexpressing mutant human amyloid
CC   beta precursor protein) or a human, suffering from chronic or acute
CC   cerebral amyloid angiopathy, preferably, Alzheimer's disease, and for
CC   ameliorating neurovascular stress comprising cerebral amyloid angiopathy
CC   in a subject, where the neurovascular stress is caused by Alzheimer's
CC   disease, aging, Down's syndrome, head trauma or stroke. This is the
CC   amino acid sequence of human receptor for advanced glycation end
CC   product (RAGE) described in the invention.
XX
SQ   Sequence   404 AA;

Query Match   7.2%; Score 147.5; DB 23; Length 404;
Best Local Similarity 22.6%; Pred No. 0.00076;
Matches   85; Conservative   40; Mismatches 114; Indels 137; Gaps 18;

Qy   15 EVIEGPQNAIVLKSSQOARFNCTVSQG--WKLIMWALSDMVVLSVRPMEPIITNDRFTS- 70
          |:|: ||| |:|:
Db 125 EIVDSASELTA--GVPNKVGTCVSEGSYPAGTLSWHLDG-------KPLVPNEKGVSV 173

Qy   71 ----QRYDQGGNFT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGSAYLTVQVMGELF 120
          :|:| ||| :|
Db 174 KEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRALRTAPIQPRVW 230

Qy 121 IP----SVNLVVAENEP-------------------CEVTCLPSHWTRLPDISWELGLLVSHSSY 162
          ||| |
Db 231 EPVPLEEVQLVV---EPEGGAVAPGGTVTLTCEVPAQPS----PQIHWMKD------- 274

Qy 163 YFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTIVIRCPQDTGGGINI 222
          ||:| ||:|
Db 275 -GVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIE-PGEEG----- 327

Qy 223 PGVLSSLPSLGFSLPTWGKVGLGLAGTMLLT------PTCTLTIRCCCRRRCGCNCC 275
          ||||| |||:|
Db 328 -----PTAGSVGGSGLGTLALALGILGGLGTRAALLGVILWQRR------ 366

Qy 276 CRCCFCCRRKRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQKITDTASLPPKS 333

Db 367 -----------------------------QRGEEKKAPENQEEEERAELN----------- 389

Qy 334 CESSDPEQRNSSCGPP 349
          :| :|| ||| |
Db 390 -QSEEPEAGESSTGGP 404

RESULT 13
ID   AAM48745 standard; protein; 404 AA.
XX
AC   AAM48745;
XX
DT   02-APR-2002 (first entry)
XX
DE   Human RAGE protein SEQ ID NO 1.
XX
KW   Human; RAGE; receptor for advanced glycated endproduct; receptor;
KW   antidiabetic; neuroprotective; cytostatic; antiinflammatory; vasotropic;
KW   nephrotropic; dermatological; antiarteriosclerotic; nootropic; diabetes;
KW   Alzheimer's disease; cancer; inflammation; kidney failure;
KW   systemic lupus; nephritis; erectile dysfunction; atherosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO200192892-A2.
XX
PD   06-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US17447.
XX
PR   30-MAY-2000; 2000US-207342P.
PR   05-MAR-2001; 2001US-0799152.
XX
PA   (TRAN-) TRANS TECH PHARMA.
XX
PI   Shahbaz M;
XX
DR   WPI; 2002-114372/15.
XX
PS   Claim 1; Fig 2; 49pp; English.
XX
CC   The invention relates to detecting receptor for advanced glycated
CC   endproducts (RAGE) modulators comprises determining the amount of RAGE
CC   protein or its fragment bound to the pre-adsorbed ligand by measuring the
CC   amount of anti-RAGE antibody bound to the solid surface. The method is
CC   useful for rapid, high-throughput identification of compounds that
CC   modulate RAGE. The compounds are useful for treating symptoms of diabetes
CC   and symptoms of diabetic late complications, amyloidoses, Alzheimer's
CC   disease, cancer, inflammation, kidney failure, systemic lupus nephritis
CC   or inflammatory lupus nephritis, erectile dysfunction and
CC   atherosclerosis.
XX
SQ   Sequence   404 AA;

Query Match   7.2%; Score 147.5; DB 23; Length 404;
Best Local Similarity 22.6%; Pred No. 0.00076;
Matches   85; Conservative   40; Mismatches 114; Indels 137; Gaps 18;

Qy   15 EVIEGPQNAIVLKSSQOARFNCTVSQG--WKLIMWALSDMVVLSVRPMEPIITNDRFTS- 70
          |:|: ||| |:|:
Db 125 EIVDSASELTA--GVPNKVGTCVSEGSYPAGTLSWHLDG-------KPLVPNEKGVSV 173

Qy   71 ----QRYDQGGNFT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGSAYLTVQVMGELF 120
          :|:| ||| :|
Db 174 KEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRALRTAPIQPRVW 230

Qy 121 IP----SVNLVVAENEP-------------------CEVTCLPSHWTRLPDISWELGLLVSHSSY 162
          ||| |

```
Db  231  EPVPLEEVQLVV---------EPEGGAVAPGGTVTLTCEVPAQPS-----PQIHWMKD---------274
Qy  163  YFVPEPSDLQOSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLIVIRCPQDTGGGINI 222
Db  275  -GVPLPLPPSPVLLIPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIE-PGEEG----327
Qy  223  PGVLSSLPGSLGFSLPTWGKVGLGLAGTMLLT------------PTCLITIRCCCCRRRCCGCNCC 275
Db  328  ------------------PTAGSVGGSGGLGTLALALGLIGLGGLTAALLIGVILWQRR----366
Qy  276  CRCCPCCRRRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQKTIDIASLPPKS 333
Db  367  ------------------QRGEERKAPENQEEEERAELN------------------389
Qy  334  CESSDPEQRNSSCGPP 349
Db  390  -QSEEPEAGESSTGGP 404
```

RESULT 14
ABB75751
ID ABB75751 standard; Protein; 592 AA.
AC ABB75751;
DT 24-JUN-2002 (first entry)
DE Human immunoglobulin superfamily member GP354.
OS Homo sapiens.
KW Human; GP354; immunoglobulin; pancreas; central nervous system;
KW diagnosis; gene therapy; pancreatitis; inflammation; tumour; cancer;
KW autoimmune disease; Alzheimer's disease; Parkinson's disease;
KW senile dementia; migraine; epilepsy; neurasthenia; neuropathy;
KW neural degeneration; antiinflammatory; cytostatic; nootropic;
KW immunosuppressive; antiparkinsonian; neuroprotective; antimigraine;
KW anticonvulsant.
FH Key Location/Qualifiers
FT Peptide   1..18
FT   /label= Signal_peptide
FT Protein   19..592
FT   /label= Mature_protein
FT Domain   19..507
FT   /note= "extracellular domain"
FT Domain   35..102
FT   /note= "Ig domain"
FT Domain   136..203
FT   /note= "Ig domain"
FT Domain   239..290
FT   /note= "Ig domain"
FT Domain   323..374
FT   /note= "Ig domain"
FT Domain   410..485
FT   /note= "Ig domain"
FT Domain   508..530
FT   /note= "transmembrane domain"
FT Domain   531..592
FT   /note= "intracellular domain"

PN WO200198360-A2.
PD 27-DEC-2001.
PF 22-JUN-2001; 2001WO-US19904.
PR 22-JUN-2000; 2000US-213611P.
PA (BIOJ/) BIOGEN INC.
PA (CARU/) CARULLI J P.
PA (LUKA/) LUKASHIN A V.
PA (KILB/) KILBURN D R.
PA (SUNC/) SUN C.
PI Carulli JP, Lukashin AV, Kilburn DR, Sun C;
DR WPI; 2002-329171/36.
DR N-PSDB; ABL53873, ABL53875.

Novel nucleic acid sequence encoding a member of immunoglobulin
superfamily, designated GP354, useful for the treatment of Alzheimer's
disease, Parkinson's disease, senile dementia, migraine and epilepsy.

Claim 17; Fig 1; 163pp; English.

CC The present sequence is the protein sequence of human GP354, a
CC novel member of the immunoglobulin superfamily. GP354 is a
CC novel pancreas-enriched integral membrane protein, which is also detected
CC at low levels in central nervous system (CNS) tissue. It shares no
CC more than 30% amino acid identity with any previously described
CC protein. The protein structure and tissue distribution of GP354
CC indicate a role in cell-cell recognition, binding, signalling and
CC adhesion events in the pancreas and CNS. The present invention
CC provides GP354 polypeptides and gp354 polynucleotides, as well as
CC vectors, host cells, antibodies and related diagnostic and
CC therapeutic methods. Claimed compositions comprising a gp354
CC nucleic acid or GP354 polypeptide are used in the treatment of
CC pancreatic injury and abnormal or disease conditions that relate to
CC the pancreas, such as acute or chronic pancreatitis, pancreatic
CC inflammation, pancreatic necrosis, exocrine insufficiency,
CC pancreatic endocrine and hormonal imbalance, pancreatic tumours and
CC associated cancers. They are also used in the treatment of
CC pancreas. CNS and abnormal or disease conditions that relate to the CNS,
CC including Alzheimer's disease, Parkinson's disease, senile dementia,
CC migraine, epilepsy, neuritis, neurasthenia, neuropathy, neural
CC degeneration and neural tumours (all claimed).
SQ Sequence 592 AA;

Query Match   6.7%; Score 138; DB 23; Length 592;
Best Local Similarity 23.8%;
Matches   55; Conservative   40; Mismatches 106; Indels   30; Gaps 10;

```
Qy  3    LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGSWKLIMWALSDMVVLSVRPMEPI 62
Db  10   LFCFRGSAGPSPHFLQQPEDLVILLGEPARLPCALGAYWGLVQWTKSGLALGGQR-----64
Qy  63   ITNDRFTSQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV-:---- 113
Db  65   ---DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYECQATOAGLRSRPAQLHVLVPPE 121
Qy  114  --QVMGELFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
Db  122  APQVLGG---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGTTFHGTLLKE 176
Qy  168  --PSDLQOSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLIVIRCPQDT 216
Db  177  GTPGSVESTILITPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVI 227
```

RESULT 15
ABB75753
ID ABB75753 standard; Protein; 594 AA.
AC ABB75753;
DT 24-JUN-2002 (first entry)
DE Human pancreas GP354.
KW Human; GP354; immunoglobulin; pancreas; central nervous system;
KW diagnosis; gene therapy; pancreatitis; inflammation; tumour; cancer;
KW autoimmune disease; Alzheimer's disease; Parkinson's disease;
KW senile dementia; migraine; epilepsy; neurasthenia; neuropathy;

```
KW   neural degeneration; antiinflamatory; cytostatic; nootropic;
KW   immunosuppressive; antiparkinsonian; neuroprotective; antimigraine;
KW   anticonvulsant.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   Peptide        1..18
FT                  /label= Signal_peptide
FT   Protein        19..594
FT                  /label= Mature_protein
XX
PN   WO200198360-A2.
XX
PD   27-DEC-2001.
XX
PF   22-JUN-2001; 2001WO-US19904.
XX
PR   22-JUN-2000; 2000US-213611P.
XX
PA   (BIOJ ) BIOGEN INC.
PA   (CARU/) CARULLI J P.
PA   (LUKA/) LUKASHIN A V.
PA   (KILB/) KILBURN D R.
PA   (SUNC/) SUN C.
XX
PI   Carulli JP, Lukashin AV, Kilburn DR, Sun C;
XX
DR   WPI; 2002-329171/36.
DR   N-PSDB; ABL53877.
XX
PT   Novel nucleic acid sequence encoding a member of immunoglobulin
PT   superfamily, designated GP354, useful for the treatment of Alzheimer's
PT   disease, Parkinson's disease, senile dementia, migraine and epilepsy -
XX
PS   Claim 17; Fig 8; 163pp; English.
XX
CC   The present sequence is the protein sequence of human pancreatic
CC   GP354, a novel member of the immunoglobulin superfamily. GP354 is
CC   a pancreas-enriched integral membrane protein, which is also
CC   detected at low levels in central nervous system (CNS) tissue. The
CC   protein structure and tissue distribution indicate a role for GP354
CC   in cell-cell recognition, binding, signalling and adhesion events
CC   in the pancreas and CNS. The invention provides GP354 polypeptides
CC   and gp354 polynucleotides, as well as vectors, host cells,
CC   antibodies and related diagnostic and therapeutic methods. Claimed
CC   compositions comprising a gp354 nucleic acid or GP354 polypeptide
CC   are used in the treatment of pancreatic injury and abnormal or
CC   disease conditions that relate to the pancreas, such as acute or
CC   chronic pancreatitis, pancreatic inflammation, pancreatic necrosis,
CC   exocrine insufficiency, pancreatic endocrine and hormonal imbalance,
CC   pancreatic tumours and associated cancers, and autoimmune disorders
CC   which affect the pancreas. They are also used in the treatment of
CC   an injury to the CNS and abnormal or disease conditions that relate
CC   to the CNS, including Alzheimer's disease, Parkinson's disease,
CC   senile dementia, migraine, epilepsy, neuritis, neurasthenia,
CC   neuropathy, neural degeneration and neural tumours (all claimed).
XX
SQ   Sequence   594 AA;

Query Match            6.6%;   Score 136;  DB 23;   Length 594;
Best Local Similarity  23.8%;  Pred. No. 0.011;
Matches  55;  Conservative  40;  Mismatches 106;   Indels  30;   Gaps  10;

Qy        3 LVIFLHGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPI 62
            |:| ||:|| ||| :||: ||||  |:| ||| :|
Db       10 LFCFPGRAGPSPHFLQQPEDLVVLGEEARLPCALGAYWGLVQWTKSGLAIGGQR----- 64

Qy       63 ITNDRFTSQRYDQGGNFTS---EMIIHNYEPSDSGNIRCSLQNSRLHG-SAYLTV----- 113
            | || |: || |||  |||     |  |  | |  || | || | || |  | |
Db       65 ---DLPGWSRYWISGNAANGQHDLHIRPVELEDEASYEQOATQAGLRSRPAQLHVLVPPE 121

Qy      114 --QVMGELFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE 167
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 16:24:32 ; Search time 37.1283 Seconds
                                    (without alignments)
                                    2142.147 Million cell updates/sec

Title: US-09-729-264-4
Perfect score: 2088
Sequence: 1 MVAGAMENRDPPGSGSGNEV..........HPQASFNLASPEKVSNTTVV 386

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 20604715 residues

Total number of hits satisfying chosen parameters: 671580

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :   SPTREMBL_21:*
                1: sp_archea:*
                2: sp_bacteria:*
                3: sp_fungi:*
                4: sp_human:*
                5: sp_invertebrate:*
                6: sp_mammal:*
                7: sp_mhc:*
                8: sp_organelle:*
                9: sp_phage:*
                10: sp_plant:*
                11: sp_rodent:*
                12: sp_virus:*
                13: sp_vertebrate:*
                14: sp_unclassified:*
                15: sp_rvirus:*
                16: sp_bacterial:*
                17: sp_archeap:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Query Match	Length	DB ID	Description
1	1692	81.0	315	4	Q9NSI5	Q9ns15 homo sapien
2	566.5	27.1	270	11	Q9D8G2	Q9d8g2 mus musculu
3	141	6.8	168	11	Q9D9I2	Q9d9i2 mus musculu
4	138	6.6	1496	4	Q92626	Q92626 homo sapien
5	136.5	6.5	173	11	Q9D4K2	Q9d4k2 mus musculu
6	135	6.5	1445	11	Q63155	Q63155 rattus norv
7	134.5	6.4	1427	13	Q91562	Q91562 xenopus lae
8	133.5	6.4	871	11	Q61987	Q61987 mus musculu
9	133.5	6.4	881	11	Q61988	Q61988 mus musculu
10	132.5	6.3	869	4	O15146	O15146 homo sapien
11	129.5	6.2	164	5	Q22048	Q22048 caenorhabdi
12	127	6.1	197	5	O17641	O17641 caenorhabdi
13	125.5	6.0	166	5	Q95QY1	Q95qy1 caenorhabdi
14	125	6.0	344	13	Q9DF61	Q9df61 gallus gall
15	125	6.0	697	11	Q9DBP0	Q9dbp0 mus musculu
16	125	6.0	697	11	Q92290	Q92290 mus musculu

RESULT 1
Q9NSI5

ID   Q9NSI5;   PRELIMINARY;   PRT;   315 AA.
AC   Q9NSI5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGSF5 protein (Fragment).
GN   IGSF5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.S.,
RA   Toyoda A., Ishii K., Totoki Y., Choi D.K., Soeda E., Ohki M.,
RA   Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.L.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RL   EMBL; AL163280; CAB90447.1; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00409; IG; 2.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
FT   NON_TER   1   1
FT   NON_TER   315   315
SQ   SEQUENCE   315 AA;   34599 MW; B69DCB23570FA3D1 CRC64;

ALIGNMENTS

17	122.5	5.9	188	5	Q18238	Q18238 caenorhabdi
18	122	5.8	6620	4	Q96AA2	Q96aa2 homo sapien
19	120.5	5.8	455	4	Q9UIRO	Q9uir0 homo sapien
20	120	5.7	313	13	O57596	O57596 gallus gall
21	120	5.7	315	13	Q9DGI5	Q9dgi5 gallus gall
22	120	5.7	344	13	O93242	O93242 gallus gall
23	120	5.7	1395	5	O44924	O44924 drosophila
24	120	5.7	1395	5	Q9W213	Q9w213 drosophila
25	119.5	5.7	6632	5	O01761	O01761 caenorhabdi
26	119	5.7	-344	11	Q99PJ0	Q99pj0 mus musculu
27	118.5	5.7	5636	4	Q96RW7	Q96rw7 homo sapien
28	118	5.7	868	11	O62838	O62838 rattus norv
29	117	5.6	868	11	Q9O94	Q9o94 homo sapien
30	116.5	5.6	364	4	Q9R094	Q9r094 mus musculu
31	116.5	5.6	868	11	Q8TD24	Q8td24 homo sapien
32	116	5.6	959	5	Q9N9Y9	Q9n9y9 drosophila
33	116	5.6	968	5	Q9W4T9	Q9w4t9 drosophila
34	116	5.6	975	5	O97174	O97174 drosophila
35	115.5	5.6	390	4	Q9H1X9	Q9h1x9 homo sapien
36	115.5	5.5	677	13	Q8QHL2	Q8qhl2 gallus gall
37	115.5	5.5	860	11	O61005	O61005 mus musculu
38	115.5	5.5	860	11	O61006	O61006 mus musculu
39	115.5	5.5	946	13	O07153	O07153 torpedo cal
40	115	5.5	1327	13	O8QHL3	O8qhl3 gallus gall
41	115	5.5	344	4	Q9P121	Q9p121 homo sapien
42	115	5.5	868	11	Q98O56	Q98o56 mus musculu
43	114.5	5.5	152	5	Q9XVX3	Q9xvx3 caenorhabdi
44	114.5	5.5	531	11	Q62319	Q62319 mus musculu
45	114.5	5.5	26926	4	Q10466	Q10466 homo sapien

```
Qy   14 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 60

Qy   74 SQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYILTVQVMGELFIPSVNLVVA 133
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 SQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYILTVQVMGELFIPSVNLVVA 120

Qy  134 ENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLT 193
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 ENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLT 180

Qy  194 CVATWKSLKARKSAATVNLTVIRCPQDTGGGTINIPGVLSSLPSLGFSLPTWGKVGLGLAGT 253
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 CVATWKSLKARKSAATVNLTVIRCPQDTGGGTINIPGVLSSLPSLGFSLPTWGKVGLGLAGT 240

Qy  254 MLLTPTCTLTIRCCCCRRRCCGCNCCCRCCFCCRRKRGFRRIQFQKKSEKEKTNKETETES 313
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 MLLTPTCTLTIRCCCCRRRCCGCNCCCRCCFCCRRKRGFRRIQFQKKSEKEKTNKETETES 300

Qy  314 GNENSGYNSDEQKTT 328
          |||||||||||||||
Db  301 GNENSGYNSDEQKTT 315
```

```
RESULT 2
ID       Q9D8G2
AC       Q9D8G2;
DT       01-JUN-2001 (TrEMBLrel. 17, Created)
DT       01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT       01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE       2010003D20Rik protein (RIKEN cDNA 2010003D20 gene).
GN       2010003D20Rik.
OS       Mus musculus (Mouse).
OC       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC       Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX       NCBI_TaxID=10090;
RN       [1]
RP       SEQUENCE FROM N.A.
RC       STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX       MEDLINE=21085660; PubMed=11217851;
RA       Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA       Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA       Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA       Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA       Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA       Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA       Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA       Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA       Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA       Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA       Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA       Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA       Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA       Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA       Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA       Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA       Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA       Hayashizaki Y.;
RL       "Functional annotation of a full-length mouse cDNA collection.";
RL       Nature 409:685-690(2001).
RN       [2]
RP       SEQUENCE FROM N.A.
RA       Strausberg R.;
RL       Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR       EMBL; AK008060; BAB25436.1; -.
DR       EMBL; BC004806; AAH04806.1; -.
MGD      MGI:1919308; 2010003D20Rik.
```

```
Query Match                 Best Local Similarity 81.0%;  Score 1692; DB 4; Length 315;
Matches 314; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Best Local Similarity 99.7%;  Pred. No. 3.6e-148;
```

```
Q9D8G2                      PRELIMINARY;       PRT;   270 AA.
```

```
Qy   14 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
          |||||||||||||||:||||||||||||||||||||||||||||||||||||||||||||
Db    1 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73

Qy   74 SQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYILTVQVMGELFIPSVNLVVA 133
          |:|  ||||:|||||||||||||||||:||||:|:||||||||||||||||||
Db   21 SGSSYQIIEGPQNVTVLKDSEAHFNCTVTHGWKLLMWTLNQMVVLSTTQGPIITNRFT 80

Qy  134 ENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLT 193
          ||||||||||:|||||||||||||||||||||||||||||||||:|:|:|||
Db   81 YASYNSTDSFISELIIHDVQPSDSGSSVQCSLLQNSHGFGSAFLSVQ------------ 125

Qy  194 CVATWKSLKARKSAATVNLTVIRCPQDTGGGTINIPGVLSSLPSLGFSLPTWGKVGLGLAGT 253
          |:||  :||||||||:||:|||||||  ||||||| |:  |:|:  ||||||
Db  126 ------------------------DSIGEEGPALPIWAILLAVAFS 148

Qy  254 MLLTPTCTLTIRCCCCRRRCCGCNCCCRCCFCCRRKRGFRRIQFQ---KKSEKEKTNK-- 307
          ||| :|| |||  |||                ||                  ||
Db  149 LLLILIIVLLIIFCCC------------------------------------------ 126
```

```
RESULT 3
ID       Q9D9I2
AC       Q9D9I2;
DT       01-JUN-2001 (TrEMBLrel. 17, Created)
DT       01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT       01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE       4931420D14Rik protein.
GN       4931420D14Rik.
OS       Mus musculus (Mouse).
OC       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC       Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX       NCBI_TaxID=10090;
RN       [1]
RP       SEQUENCE FROM N.A.
RC       STRAIN=C57BL/6J; TISSUE=TESTIS;
RX       MEDLINE=21085660; PubMed=11217851;
RA       Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA       Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA       Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA       Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA       Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA       Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA       Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA       Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA       Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA       Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA       Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA       Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA       Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA       Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA       Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA       Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA       Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
```

```
Query Match                 Best Local Similarity 27.1%;  Score 566.5; DB 11; Length 270;
Matches 131; Conservative 42; Mismatches 71; Indels 135; Gaps 5;
Best Local Similarity 34.6%;  Pred. No. 3.3e-44;
```

```
DR       InterPro; IPR003599; Ig.
DR       InterPro; IPR003006; Ig_MHC.
DR       SMART; SM00409; IG;
DR       PROSITE; PS00290; IG_MHC; UNKNOWN_1.
SQ       SEQUENCE 270 AA; 29604 MW; A39C273DAA950DFE CRC64;
```

```
Q9D9I2                      PRELIMINARY;       PRT;   168 AA.
```

```
Qy  194 CVATWKSLKARKSAATVNLTVIRCPQDTGGGTINIPGVLSSLPSLGFSLPTWGKVGLGLAGT 253
          |:||||||||||||||||||||||||||||| |||||: |:||:||||||||||
Db  194 ETKLKSGKENYGYSSDEAKAAQTASLPPKSAEVSLPEKRSSSL--PYOELNKHQPGPATH 251

Qy  254 MLLTPTCTLTIRCCCCRRRCCGCNCCCRCCFCCRRKRGFRRIQFQ---KKSEKEKTNK--- 307
          ||| :|| ||| |||                 ||                   :|:
Db  252 PRVSFDIASPQKVRNVTLV 270
```

```
Qy  308 ETETESGNENSGYNSDEQKTTDTASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASH 367
          |:|:|:|:|||||||||||:||   ||
Db  194 ETKLKSGKENYGYSSDEAKAAQTASLPPKSAEVSLPEKRSSSL--PYOELNKHQPGPATH 251

Qy  368 PQASENIASPEKVSNTTVV 386
          |:|:||:||||:|:||| ||:|
Db  252 PRVSFDIASPQKVRNVTLV 270
```

RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK008992; BAB24782.1; -.
DR   MGD; MGI:1913992; 4931420D14Rik.
SQ   SEQUENCE   168 AA; 18931 MW; 7A2BD279612A5E94 CRC64;

Qy   Best Match
     Best Local Similarity   32.0%; Pred. No.3.8e-05;
Matches 39; Conservative 20; Mismatches 49; Indels 14; Gaps 6;

Qy   260 CTLTLTRCCCCRRRCCGCNC--CCRCCFCCCRRKRGFR----IQFQKKSEKEKTNKETETESG 314
              |:|:| ||||| :: : | : ||: | |: ||: ||| : | ||:||| | |
Db   55 CSLSRSCCCCCRCRCC-CHCRCCRC--CCSRSRFRSRRTTLKFFQITEKGEQSLQRRIRRQ 111

Qy   315 NENSGYNSDEQKTDTASLPPKSCCESSDDPEQRNSSCGPPHQRADQRPPRPASHPQASFNL 374
              |:|: | | | : | || :: :|| || || | :| :|::|| |: :| |: |
Db   112 LTRSQLELIEPE--PTMALEPSEITTVAFFSHKNANVSDP----EEVPCLDSDPFPNGDL 165

Qy   375 AS 376
              ||
Db   166 AS 167

RESULT 4
ID   Q92626   PRELIMINARY;   PRT;   1496 AA.
AC   Q92626;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   MYELOBLAST KIAA0230 (Fragment).
GN   KIAA0230.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O.,
RA   Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RA   Mitchell M.S., Kan-Mitchell J., Minev B., Edman C., Deans R.J.;
RT   "Identification of a novel melanoma gene (MG50) - likely the gene for
RT   IL-1 receptor antagonist - which encodes epitopes recognized by human
RT   cytolytic T lymphocytes.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RA   Trent J.M.;
RT   "Assignment of a human melanoma associated gene MG50 (D2S448) to
RT   chromosome 2p25.3 by fluorescence in situ hybridization.";
RL   Genomics 22:243-244(1994).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MELANOMA;
RX   MEDLINE=95048383; PubMed=7959781;
RA   Weiler S.R., Taylor S.M., Deans R.J., Kan-Mitchell J., Mitchell M.S.,
RL   TISSUE=BONE MARROW;
DR   EMBL; D86983; BAA13219.1; -.
DR   EMBL; AF200348; AAF06354.1; -.
DR   HSSP; P05164; 1CXP.
DR   InterPro; IPR002007; Anim_peroxidase.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR000483; LRR_Cterm.
DR   InterPro; IPR000372; LRR_Nterm.

DR   InterPro; IPR003591; LRR_typ.
DR   InterPro; IPR0011007; VWF_C.
DR   Pfam; PF03098; An_peroxidase; 1.
DR   Pfam; PF00047; Ig; 4.
DR   Pfam; PF00560; LRR; 5.
DR   Pfam; PF01463; LRRCT; 1.
DR   Pfam; PF00093; VWC; 1.
DR   PRINTS; PR00457; ANPEROXIDASE.
DR   SMART; SM00408; IGc2; 4.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 4.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; UNKNOWN_1.
KW   Immunoglobulin domain.
FT   NON_TER   1   1
FT   NON_TER   1   1
SQ   SEQUENCE   1496 AA; 167209 MW; E9B9A7069BF1ABFF CRC64;

Qy   Query Match
     Best Local Similarity   25.1%; Pred. No. 0.0013;
Matches 78; Conservative 38; Mismatches 127; Indels 68; Gaps 17;

Qy   2 VAGAMENRDPP-----GSGSGNEVIEGPQNARVLKGSQARFNCTVS-QGWKLIMWALSDMV 56
              | ::: :| : |: |: | :| :|| |:| : : ||:| || : : | ||:
Db   339 VAGEVKTQEVTLRYFGSPARPTFVIQPQNIEVLVGESVTLECSATGHPPPRISWTRGDRT 398

Qy   57 VLSVRPMEPIITNDRFTSQRIDQGGNFTSEMIIHNVEPSDSGNIRCSLQNS--RLHGSAY 114
              |: |: |: || :   :: :: : :| :|: || ||| || || | :|: ||
Db   399 PLFVDPRVNITPS-----------GG----LYIQNVVQGDSGEYACSATNNIDSVHATAF 443

Qy   115 LTVQVMGELFIPSVNLLVVAENE----PCEVTCLPSIHWTLPDISW-ELGLIVSHSSYFPV 169
              | :| | || : | : ::|: | ::| | : |: | || :||| |:|:| : |
Db   444 IIVQALPQFTVTPQDRVVIEGQTVDFQCEAKGNPP-----PVIAWTKGGSQLSVDRRHLV 498

Qy   170 PEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV----IRCPQDT- 220
              | : | | | | | :| || | |:| ||: |:|: | ||| |:
Db   499 LSSGTLR--ISGVALHDQ--GQYECQAV--NIIGSQKVVAHLTVQPRVTPVFFASIPSDTT 552

Qy   221 ---GGGINIPGVLSSLPSLGFSLP--TWGKVGLGL--AGTMLLTPTCTLTI---------- 264
              | : :| |: :| |: | | |:|:| : | : : :   :| :
Db   553 VEVGANVQLP-----CSSQGEPEPAITWNKDGVQVTESGKFHISPEGGFLTINDVGPADAG 607

Qy   265 RCCCCRRRCCG 275
              | | :| |
Db   608 RYECVARNIIG 618

RESULT 5
ID   Q9D4K2   PRELIMINARY;   PRT;   173 AA.
AC   Q9D4K2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   4931420D14Rik protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA　Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA　Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA　Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA　Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA　Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA　Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA　Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA　Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA　Hayashizaki Y.;
RT　"Functional annotation of a full-length mouse cDNA collection.";
RL　Nature 409:685-690(2001).
DR　EMBL; AK016467; BAB30253.1; -.
DR　MGD; MGI:1913992; 4931420D14Rik.
SQ　SEQUENCE　173 AA;　19581 MW;　147B6F155AC29FDF CRC64;

Qy　260 CTLFIRCCCCRRRCCGCNC-CCRCRCPFCCRRKRGRFIQFQKKSEKEKTNEETESGNENS 318
　　　　|||||.|||||||||||| |||||.||||.|.|||||||.|||||||||||||||||
Db　55　CSLSRSCCCCRCRCC-CYCRCCRC--CCSRSRFRSRTLRVRDPKFFQ--ITEKGEQSL 109

Qy　319 GYNSDEQKTTD-------TASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQ 369
　　　　||||||| ||.|||||| |||||||||||||| ||||||||||||||.||||||||
Db　110 QRRIRQLTRSQLELIEPEPTMALEPSEITVAFFSHKNANVSDP----EEVPCLDSDPF 165

Qy　370 ASFNIAS 376
　　　　|:|||||
Db　166 PNGDLAS 172

Query Match　　　　6.5%;
Best Local Similarity　31.5%;
Matches　40;　Conservative　16;　Mismatches　52;　Indels　19;　Gaps　6;

RESULT 6
Q63155
ID　Q63155;
AC　Q63155;
DT　01-NOV-1996 (TrEMBLrel. 01, Created)
DT　01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT　01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE　Colorectal tumor suppressor.
DCC.
GN　DCC.
OS　Rattus norvegicus (Rat).
OC　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC　Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX　NCBI_TaxID=10116;
RN　[1]
RP　SEQUENCE FROM N.A.
RX　MEDLINE=97015074; PubMed=8861902;
RA　Keino-Masu K., Masu M., Hinck L., Leonardo E.D., Chan S.S.Y.,
RA　Culotti J.G., Tessier-Lavigne M.;
RT　"Deleted in Colorectal Cancer (DCC) encodes a netrin receptor.";
RL　Cell 87:175-185(1996).
RN　[2]
RP　SEQUENCE OF 387-420 FROM N.A.
RX　MEDLINE=90100559; PubMed=2294591;
RA　Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA　Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA　Vogelstein B.;
RT　"Identification of a chromosome 18q gene that is altered in colorectal
RT　cancers.";
RL　Science 247:49-56(1990).
DR　EMBL; U68725; AAB41099.1; -.
DR　EMBL; M32291; AAA41086.1; -.
HSSP; P56276; 1TLK.
DR　InterPro; IPR003962; FnIII_repeat.
DR　InterPro; IPR003961; FN_III.
DR　InterPro; IPR003598; Ig-c2.
DR　InterPro; IPR003600; Ig_like.
DR　InterPro; IPR003006; Ig_MHC.
DR　Pfam; PF00041; fn3; 6.
DR　Pfam; PF00047; ig; 4.
DR　PRINTS; PR00014; FNTYPEIII.
SMART; SM00060; FN3; 6.

DR　SMART; SM00408; IGc2; 3.
DR　SMART; SM00410; IG_like; 2.
KW　Immunoglobulin domain; Repeat.
SQ　SEQUENCE　1445 AA;　157940 MW;　084F62595448I988 CRC64;

Query Match　　　　6.5%;
Best Local Similarity　25.5%;　Score 135;　DB 11;　Length 1445;
Matches　56;　Conservative　32;　Mismatches　90;　Indels　42;　Gaps　9;

Qy　9　RDPPGGSGNEV----------------------IEGPQNARVLKGSQARFNCTVSQGW--KLIMW 50
　　　　|:|: :|||| 　　　　　　　　　　 ||| ||||||||||||||||||| :||||
Db　215 RNPASTRTGNEAEVRILSDPGLHRQLYFLQRPSNVIAIEGKDAVLECCVS-GYPPSFTW 273

Qy　51　ALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSR 108
　　　　|:|| ||| |:||| | :|:|||||||||||||||||||||||:||||| ||||||
Db　274 LRGEEVI------------------QLRSKKYSLLGG--SNLLISNVTDDDSGTYTCVVTYKNEN 318

Qy　109 LHGSAYLTVQVMGELFIPSVNLVVAENPECEVTCLPSHWTWLPDISW-ELGLLVSHSSYY 167
　　　　|||:|||||| ||||: |:||||||||||: :|||:|||||||||| :|||:||:||||
Db　319 ISASAELTVLVPPWFLNHPSNLIAYESMDIEPECAVSGKP-VPTVNWMKNGDVVIPSDYF 377

Qy　168 FVPEPSDLQSAVSILALIPQSNGTLLTCVATWKSLKARKSA 207
　　　　|:|| |:| :|||:||||:|||| |||||:|||||:|||
Db　378 QIVGGSNLR---ILGVVKSDEGFYQCVAENEAGNAQSSA 413

RESULT 7
Q91562
ID　Q91562;
AC　Q91562;
DT　01-NOV-1996 (TrEMBLrel. 01, Created)
DT　01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT　01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE　Tumor suppressor.
XDCCA.
GN　XDCCA.
OS　Xenopus laevis (African clawed frog).
OC　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC　Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC　Xenopodinae; Xenopus.
OX　NCBI_TaxID=8355;
RN　[1]
RP　SEQUENCE FROM N.A.
RX　MEDLINE=95113183; PubMed=7813784;
RA　Pierceall W.E., Reale M.A., Candia A.F., Wright C.V., Cho K.R.,
RA　Fearon E.R.;
RT　"Expression of a homologue of the deleted in colorectal cancer (DCC)
RT　gene in the nervous system of developing Xenopus embryos.";
RL　Dev. Biol. 166:654-665(1994).
DR　EMBL; U10986; AAA70168.1; -.
HSSP; P40189; 1BQU.
DR　InterPro; IPR003962; FnIII_repeat.
DR　InterPro; IPR003961; FN_III.
DR　InterPro; IPR003598; Ig-c2.
DR　InterPro; IPR003600; Ig_like.
DR　InterPro; IPR003006; Ig_MHC.
DR　Pfam; PF00041; fn3; 6.
DR　Pfam; PF00047; ig; 4.
DR　PRINTS; PR00014; FNTYPEIII.
DR　SMART; SM00060; FN3; 6.
DR　SMART; SM00408; IGc2; 3.
DR　SMART; SM00410; IG_like; 2.
KW　Immunoglobulin domain; Repeat.
SQ　SEQUENCE　1427 AA;　156533 MW;　61FEA12C8A674972 CRC64;

Query Match　　　　6.4%;
Best Local Similarity　25.4%;　Score 134.5;　DB 13;　Length 1427;
Matches　65;　Conservative　33;　Mismatches　93;　Indels　65;　Gaps　15;

Qy　12　PGSGS-GNEV------------------IEGPQNARVLKGSQARFNCTVSQGW--KLIMWAL 52
　　　　||| ||| 　　　　　　　　 ||| ||||||||||||||||||||| :||||||
Db　217 PGSARVGNEAELRILSEGGLHRQQVFLQRPSNVVAIEGQDAVLECAVS-GYPTPTIVWMQ 275

Qy　53　SDMVVLSVRPMEPIITNDRFTSQRYDQGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLH 110

RESULT 8

```
Qy  276 GD-------------EPVPIRTR----KYSVLGG--SNLLISNVTDDDAGAYTCVATYKNENTS 320
Db  111 GSAYLTVQVMGELFIPSVNLLVVAENEPCEVTCL----PSHWTWLPDISW-ELGLLVSHSS 165
Qy  321 FSADLTVMVPPQFLNHPANLYAYESMDIEFECAVSGKPS----PTVWKMTKNGEVVIPSD 375
Db  166 YYFVPEPSDLQSAVSILALTPQSNGTLTCVA--------TWKSLKARKSATVNLTVI-RC 216
Qy  376 YFQIVDGSNLR----ILGLVKSDEGYYQCIAENEAGNIQTYAQLIIPDDAVPSSSILPSA 431
Db  217 PQDTGGGINIPGVLSS 232
Qy  432 PRDV------VPVLVSS 442
```

```
ID   Q61987     PRELIMINARY;    PRT;   871 AA.
AC   Q61987;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Nsk2 protein precursor.
GN   MUSK OR NSK2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MYOBLAST;
RX   MEDLINE=95349951; PubMed=7624144;
RA   Ganju P., Walls E., Brennan J., Reith A.D.;
RT   "Cloning and developmental expression of Nsk2, a novel receptor
RT   tyrosine kinase implicated in skeletal myogenesis,.";
RL   Oncogene 11:281-290(1995).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; X86444; CAA60165.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:103581; Musk.
DR   InterPro; IPR002453; Beta_tubulin.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000024; Fz_domain.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF01392; Fz; 1.
DR   Pfam; PF00047; Ig; 3.
DR   Pfam; PF00069; Pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   SMART; SM00038; FZ; 1.
DR   PROSITE; PS50038; FZ; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
KW   ATP-binding; Immunoglobulin domain; Serine/threonine-protein kinase;
KW   Signal; Transferase.
FT   SIGNAL        1     21       POTENTIAL.
SQ   SEQUENCE   871 AA;  97047 MW;  F3C53DC6AFE702AB CRC64;
```

```
Query Match             6.4%;
Best Local Similarity   21.6%;
Matches   73;  Conservative  45;  Mismatches  127;  Indels  93;  Gaps  17;
Score 133.5;  DB 11;  Length 871;  Pred. No. 0.0016;

Qy      4 GAMENRDPPGSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRP 62
```

RESULT 9

```
Qy  113 GALQVRMKP-----------KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD------------- 158
Db   63 MEPIIINDRFTSQRYDQGGNFTSEMIIHNYEPSDSGNIRCSLQNSRLHGSAY-----LTV 117
Qy  159 -NALRENSRIAALE-------SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKLVKLEV 208
Db  118 QVMGELFIPSVNLLVVAENEPCEVTC-----LPSHWTWLPDISW-ELGLLVSHSSYYFVPEP 172
Qy  209 EVIGRILRAPESHNVTFGSFVTLRCTEIGIP----VPTISWIENGNAVSSGSIQESVKD 263
Db  173 SDLQSAVSILALTPQSNGTLTCVAT----WKSLKARKSATVNLTVIR--------CPQD 219
Qy  264 RVIDSRLQLFITKP---GLYTCIATNKHGEKFSTAKAAATVSIAEWSKSQKDSQGYCAQY 320
Db  220 TGGGINIPG------VLSSLPSLGFSLP--------------TWGKVGLGLAGTIMLLTPTCTLTI 373
Qy  321 RGEGVLMQGPGEKMLLVFLPTTSHRDPEDAQELLIHTAWNEL--------KAVSPLCRPAA 373
Db  265 RCCCCRRRCCGCN----------CCCRCC-------FCCR 287
Qy  374 EALLCYHLFLECSPGVVPTPMPICREYCLAVKELFCAK 411
```

```
ID   Q61988     PRELIMINARY;    PRT;   881 AA.
AC   Q61988;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Nsk2 protein precursor.
GN   MUSK OR NSK2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MYOBLAST;
RX   MEDLINE=95349951; PubMed=7624144;
RA   Ganju P., Walls E., Brennan J., Reith A.D.;
RT   "Cloning and developmental expression of Nsk2, a novel receptor
RT   tyrosine kinase implicated in skeletal myogenesis,.";
RL   Oncogene 11:281-290(1995).
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR   EMBL; X86445; CAA60166.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:103581; Musk.
DR   InterPro; IPR002453; Beta_tubulin.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000024; Fz_domain.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF01392; Fz; 1.
DR   Pfam; PF00047; Ig; 3.
DR   Pfam; PF00069; Pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00219; TyrKc; 1.
DR   SMART; SM00038; FZ; 1.
DR   PROSITE; PS50038; FZ; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
KW   ATP-binding; Immunoglobulin domain; Serine/threonine-protein kinase;
KW   Signal; Transferase.
```

FT   SIGNAL
SQ   SEQUENCE   881 AA;   98435 MW;   POTENTIAL.
                                        EA0D0282E9B28ED7 CRC64;

Query Match
Best Local Similarity   21.6%;   Score 133.5;   DB 11;   Length 881;
Matches 73;   Conservative 45;   Mismatches 127;   Indels 93;   Gaps 17;

Qy   4 GAMENRDPPGSGSGNEVIEGPQONARVLKGSQARFNCTVSQGWK-LIMWALSDMVLSVRP 62
           ||||: ::|:: ||::::|:|:|: ::: |:|| |: |:|::
Db   113 GALQVKMKP-------KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD-------- 158

Qy   63 MEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPDSDSGNIRCSLQNSRLHGSAY------LTV 117
           ::|||: :: |: ||:|| :::|:||: :|: ::|: |:: :|::|::
Db   159 -NALRENSRIAALE------SGSLRIHNVOKEDAGQYRCVAKNSL--GTAYSKLVKLEV 208

Qy   118 QVMGELFIPSVNLVVAAENEPCEVTC----------LPSHWTWLPDISW-ELGLLVSHSSYYFVPEP 172
           :||: |:|||: |: ||| || ||
Db   209 EVLGRILRAPESHNVTFGSFVTLRCTEIGIP-----VPTISWIENGNAVSSGSIQESVKD 263

Qy   173 SDLQSAVSILALTPQSNGTLTCVAT----WKSLKARKSATVNLTVIR--------CPQD 219
           |:|| |: :|:|:|: |::|
Db   264 RVIDSRLQLFITKP----GLYCIATNKHGEKFSTAKAAATVSIAEWSKSQKDSQGYCAQY 320

Qy   220 TGGGINIPG-----VLSSLPSLGFSLP-------------TWGKVGLGLAGTMLLTPTCLTI 264
           :| ::|
Db   321 RGEGVLMQGPGEKMLLVFLPTTSHRDPEDAQELLIHTAWNEL------KAVSPLCRPAA 373

Qy   265 RCCCCRRRCCGCN---------CCRCC---------FCCR 287
Db   374 EALLCYHLFLECSPGVVPTPMPICREYCLAVKELFCAK 411

DR   SMART; SM00408; IGC2; 3.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS50038; FZ; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
KW   ATP-binding; Immunoglobulin domain; Kinase; Receptor; Transferase;
KW   Tyrosine-protein kinase.
SQ   SEQUENCE 869 AA; 97056 MW; 3DDC20E179FA010C CRC64;

RESULT 10
ID   O15146
AC   O15146;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   "Receptor tyrosine kinase specific for the skeletal muscle lineage:
     expression in embryonic muscle, at the neuromuscular junction, and
     after injury."
GN   MUSK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96009854; PubMed=7546737;
RA   Valenzuela D.M., Rojas E., Yancopoulos G.D.;
RA   Compton D.L., Nunez L., Park J.S., Stitt T.N., Distefano P.S.,
RA   LeBeau M.M., Fernald A.A., Copeland N.G., Jenkins N.A., Burden S.J.,
RA   Glass D.J., Yancopoulos G.D.;
RL   Neuron 15:573-584(1995);
RN   [2]
RL   SEQUENCE FROM N.A.
RA   Valenzuela D.M., Rojas E., Yancopoulos G.D.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
HSSP; P11362; IFGK.
EMBL; AF006464; AAB63044.1; -.
DR   InterPro; IPR002453; Beta_tubulin.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR000024; Fz_domain.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF01392; Fz; 1.
DR   Pfam; PF00047; ig; 3.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.

Query Match
Best Local Similarity   20.8%;   Score 132.5;   DB 4;   Length 869;
Matches 89;   Conservative 55;   Mismatches 178;   Indels 105;   Gaps 20;

Qy   4 GAMENRDPPGSGSGNEVIEGPQONARVLKGSQARFNCTVSQGWK-LIMWALSDMVLSVRP 62
Db   113 GALQVKMKP-------KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD-------- 158

Qy   63 MEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPDSDSGNIRCSLQNSRLHGSAY------LTVQVMG 121
Db   159 -SPLRENSRIAVLE------SGSLRIHNVOKEDAGQYRCVAKNSL--GTAYSKVKLEV 208

Qy   122 ELFIPSVNLVVAENEP----------CEVTCLPSHWTWLPDISW-ELGLLVSHSSYYFVPEP 172
Db   209 EVFARILRAPESHNVTFGSFVTLHCTATGIP-----VPTITWIENGNAVSSGSIQESVKD 263

Qy   173 SDLQSAVSILALTPQSNGTLTCVAT----WKSLKARKSATVNLTVIRCPDTGGG 320
Db   264 RVIDSRLQLFITKP----GLYCIATNKHGEKFSTAKAAATVSIAEWSKPQKDNKGYCAQY 320

Qy   224 -----INIPGVLSSLPSLGFSL--------PTWGKVGLGLAGTMLLTPTCLTLTIRCC 267
Db   321 RGEVCNAVLAKDALVFLNTSYADPEEAQELLVHTAWNEL------KVVSPVCRPAAEAL 373

Qy   268 CCRRRCCGCN--------CCRCC------FCCRRKRGFRIOFQKKSEKETKNKETETE 312
Db   374 LCNHIFQECSPGVVPTPIPICREYCLAVKELFCAKE----WLVMEEKTHRGLYRSEMHLL 429

Qy   313 SGNENSGGYNSDEQKTTDTASLPPKSCESSDPEQRNSSCGPPHQRADQRP----PRPASHP 368
Db   430 SVPECSKLPSMHWDPTACARLP-----HLDYNKENLKTFPP--MTSSKPSVDIPNLPSSS 482

Qy   369 QASFNLA 375
Db   483 SSSFSVS 489

RESULT 11
ID   Q22048
AC   Q22048;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   T01B7.8 protein.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sims M.A.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
RL   EMBL; Z66499; CAA91301.1; -.

PRELIMINARY;   PRT;   869 AA.

PRELIMINARY;   PRT;   164 AA.

```
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001271; Defensin_mammal.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001007; VWF_C.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; UNKNOWN_1.
DR   PROSITE; PS00269; DEFENSIN; UNKNOWN_1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01208; VWFC; UNKNOWN_1.
SQ   SEQUENCE 164 AA; 16499 MW; C002D48D36C9FCED CRC64;

Qy   157 LGILVSHSSYFVPEPSDLQSAVSTIALIPQSNGTLTCVATWKSLKARKSATVNLTVIRC 216
             ::: ::::   :::::   :  :::::: :  :::   :::::::   :::
Db     6 LAILLAIGTFIAV--SQVQSAV---LPVSSTELATVGTDVSTASTAIDTLGNSSSRV 57

Qy   217 PQDTGGGINIPGVLSSLPSLGFSLPTWGKVLGILAGTMLLTPTCTLTIRCCCCRRRCCGC 276
           :  :  :       :::   : ::   ::: :::   : : :   :::  :
Db    58 KRQGCGCGCCGC--------------GCGCGCGGGGG--CGCCCCRPRCCCCCRRCCTC 101

Qy   277 --NCCC-RCCPCCR 287
           ::::  ::: :::
Db   102 CRTCCCTRCCTCCR 115
```

RESULT 12
```
ID   Q17641                     PRELIMINARY;      PRT;   197 AA.
AC   Q17641;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical 20.6 kDa protein.
GN   C04G6.7.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Anderson K., Chissoe S.;
RT   "The sequence of C. elegans cosmid C04G6.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   "Direct Submission.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U55854; AAK68161.1; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001271; Defensin_mammal.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001007; VWF_C.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; UNKNOWN_1.
DR   PROSITE; PS00269; DEFENSIN; UNKNOWN_1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01208; VWFC; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE 197 AA; 20596 MW; FB5F9457BF99B8AD CRC64;

Query Match        6.1%; Score 127; DB 5; Length 197;
```

Best Local Similarity   32.1%; Pred. No. 0.00043; DB 5; Length 164;
Matches 43; Conservative 11; Mismatches 53; Indels 27; Gaps 6;

RESULT 13
```
ID   Q95QY1                     PRELIMINARY;      PRT;   166 AA.
AC   Q95QY1;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical 17.0 kDa protein.
GN   C04G6.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   "Direct Submission.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Anderson K., Chissoe S.;
RT   "The sequence of C. elegans cosmid C04G6.";
RL   Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U55854; AAK68158.1; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001271; Defensin_mammal.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001007; VWF_C.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; UNKNOWN_1.
DR   PROSITE; PS00269; DEFENSIN; UNKNOWN_1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01208; VWFC; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE 166 AA; 16971 MW; 9D9D130351BB50F1 CRC64;

Query Match        6.0%; Score 125.5; DB 5; Length 166;
```

Best Local Similarity   29.6%; Pred. No. 0.00093;
Matches 32; Conservative   8; Mismatches 42; Indels 26; Gaps 4;

```
Qy   260 CTLTIRCCCCRRRCCGC--NCCC-RCCPCCRR------------------KRGFRIQFQK 298
            :: :::: :: ::::   :::: ::: :::                  :  :
Db    87 CCCRPRCCCCCRRCCTCCRCRTCCCTRCCTCCRPCCCGCGCGGCGGCGGCGGGGGGHFLKSTAMARPK 146

Qy   299 KSEKEKTNKETETESGNENSGYNSDEQKTTDTASLPPKSCESSDPEQR 346
           :   :                             ::::    :: ::
Db   147 PLIEEKAKKQENDKKNREASRN-----CCDLASVEVRRPNFADYNYR 189
```

RESULT 14
```
ID   Q9DF61                     PRELIMINARY;      PRT;   344 AA.
AC   Q9DF61;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

Query Match        6.0%; Score 125.5; DB 5; Length 166;
```

Best Local Similarity   64.5%; Pred. No. 0.001;
Matches 20; Conservative   0; Mismatches   8; Indels   3; Gaps   2;

```
Qy   260 CTLTIRCCCCRRRCCGC--NCCC-RCCPCCRR 287
            :: :::: :: ::::   :::: ::: :::
Db    86 CCCRPRCCCCCRRCCTCCRCRTCCCTRCCTCCR 116
```

```
DE   OBCAM alpha 1 isoform.
GN   OBCAM.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20499204; PubMed=11042360;
RA   Lodge A.P., Howard M.R., McNamee C.J., Moss D.J.;
RT   "Co-localisation, heterophilic interactions and regulated expression
RT   of IgLON family proteins in the chick nervous system.";
RL   Brain Res. Mol. Brain Res. 82:84-94(2000).
DR   EMBL; AF292934; AAG01877.1; -.
DR   InterPro; IPR003599; IG.
DR   InterPro; IPR003598; IG_c2.
DR   InterPro; IPR003600; IG_like.
DR   InterPro; IPR003006; IG_MHC.
DR   Pfam; PF00047; IG; 3.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain.
SQ   SEQUENCE 344 AA; 37531 MW; 37FE6051CBF0E7B4 CRC64;

Query Match          6.0%;  Score 125; DB 13;  Length 344;
Best Local Similarity 25.4%;  Pred. No. 0.0029;
Matches   64;  Conservative   33;  Mismatches   97;  Indels   58;  Gaps   14;

Qy    26 NARVLKGSQARFNCTVSQGWKLIMWALSDMVVL------SVRPMEPIITNDRFTSQRYD 78
                :|||:       ::                       |:|:|
Db    44 NVTVRQGESATLRCTVDDRVRRVAW-LNRSTILYAGNDKWSIDNRVVILSN---TKTQY- 98

Qy    79 QGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSRLHGSAYLTVQVMGELFIPSVNLVV 132
            :  |||| |  :: :                  :|:|
Db    99 -----SIKIHNVDVDEGPYTCSVQTDNHPKTSRVH---LIVQVPPQIVNISSDITV 147

Qy   133 AENEPCEVTCL-----PSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQS 188
                 | :|  :|      |:       :|
Db   148 NEGSSVTIMCLAFGRPE----PTVTWR--HLSGKQGGFVSEDEYLE----ITGITREQ 195

Qy   189 NGTLTCVATWKSLKARKSATVNLIV------IRCPODTGGGINIPGVL----SSLPSLGFS 239
            | | |
Db   196 SGEYECSAV-NDVAVPDVRKVKVTVNYPPYISNAKNTGASVGQKGILQCEASAVPVAEFQ 254

Qy   240 LPTWGKVGLGLA 251
            | ||
Db   255 ---WFKEDTRLA 263
```

```
RESULT 15
Q9DBP0

ID   Q9DBP0        PRELIMINARY;     PRT;    697 AA.
AC   Q9DBP0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Solute carrier family 34 (sodium phosphate), member 2.
GN   SLC34A2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LUNG;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK004832; BAB23600.1; -.
DR   MGD; MGI:1342284; Slc34a2.
DR   InterPro; IPR003841; Na_Pi_cotrans.
DR   Pfam; PF02690; Na_Pi_cotrans; 1.
DR   TIGRFAMs; TIGR01013; 2a58; 1.
SQ   SEQUENCE 697 AA; 76244 MW; 2A7B9384857EF16F CRC64;

Query Match          6.0%;  Score 125; DB 11;  Length 697;
Best Local Similarity 25.8%;  Pred. No. 0.0074;
Matches   41;  Conservative   12;  Mismatches   45;  Indels   61;  Gaps   8;

Qy   233 LPSLGFSLPTWGKVGLGLAGTMLLTPTCLTLTIRCCCCRRCCG-----CNC-CCRCCFCC 286
            || ::|:|:|  ::|          :|  :||||| | :|||       ||  ||:| ||
Db   593 LPLWMHSLKPWDNV------ISLATTC-FQRRCCCCCRVCCCRVCCMVCGCKCCRCSKCC 644

Qy   287 RRKRGFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKTTDFASLPPKSCESSSDPEQR 346
            |                                 :|  :|::|
Db   645 R-------------DQGEE------EEKEQD----IPVKASGAFDNAAM 671

Qy   347 NSSCGPPHQRADQRPPRPASHPQASFNLASPEKVSNTTV 385
                                      :      |:||||
Db   672 SKEC-----------QDEGKGQVEVLSMKALSNTTV 696

Search completed: April 28, 2003, 21:07:38
Job time : 41.1283 secs
```

OM protein - protein search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.4_p5_4578

Run on: April 28, 2003, 16:24:30 ; Search time 10.3692 seconds
(without alignments)
1543.990 Million cell updates/sec

Title: US-09-729-264-4
Perfect score: 2088
Sequence: 1 MVAGAMENRDPPGSGSGNEV.........HPQASFNLASPEKVSNTTVV 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	145.5	7.0	404	1	Q15109	RAGE_HUMAN	homo sapien
2	136	6.5	1447	1	P43146	DCC_HUMAN	homo sapien
3	132	6.3	1447	1	P70211	DCC_MOUSE	mus musculu
4	126	6.0	1914	1	Q15746	KMLS_HUMAN	homo sapien
5	126	6.0	337	1	Q98892	G5SA_CHICK	gallus gall
6	125	5.8	353	1	Q9073	CEPU_CHICK	gallus gall
7	122	5.8	344	1	Q62718	NTR1_RAT	rattus norv
8	121	5.8	1070	1	Q13308	PTK7_HUMAN	homo sapien
9	119	5.7	862	1	P35329	CD22_MOUSE	mus musculu
10	114.5	5.5	862	1	Q92859	NEO1_HUMAN	homo sapien
11	113	5.5	1461	1	Q14982	OPCM_HUMAN	homo sapien
11	112.5	5.4	345	1	Q28173	RAGE_BOVIN	bos taurus
12	112.5	5.4	416	1	P1834	OPCM_BOVIN	bos taurus
13	111.5	5.3	345	1	Q90610	NEO1_CHICK	gallus gall
14	111.5	5.3	1443	1	Q92154	SMP_COTJA	coturnix co
15	111	5.3	620	1	P97603	NEO1_RAT	rattus norv
16	111	5.3	1377	1	P32736	OPCM_RAT	rattus norv
17	110.5	5.3	345	1	P98160	PGBM_HUMAN	homo sapien
18	110.5	5.3	4393	1	Q03751	CSP_DROME	drosophila
19	109.5	5.2	249	1	P20138	CD33_HUMAN	homo sapien
20	109.5	5.2	364	1	Q06561	UN52_CAEEL	caenorhabdi
21	109	5.2	2481	1	Q62556	BUTY_MOUSE	mus musculu
22	106.5	5.1	524	1	P36335	NCA2_XENLA	xenopus lae
23	106	5.1	1092	1	Q99795	A33_HUMAN	homo sapien
24	104	5.0	319	1	Q08180	ICCR_DROME	drosophila
25	104	5.0	764	1	P97792	CXAR_MOUSE	mus musculu
26	103.5	5.0	365	1	Q05793	PGBM_MOUSE	mus musculu
27	103.5	5.0	3707	1	P15364	AMAL_DROME	drosophila
28	102.5	4.9	333	1	Q62786	FPRP_RAT	rattus norv
29	102.5	4.9	879	1	Q06418	TYO3_HUMAN	homo sapien
30	102.5	4.9	890	1	P78310	CXAR_HUMAN	homo sapien
31	102	4.9	365	1	Q2OJ8	KILO_RAT	rattus norv
32	100.5	4.8	348	1	Q24372	LACH_DROME	drosophila
33	99	4.7	359	1			
34	99	4.7	830	1	SREC_HUMAN	homo sapien
35	98	4.7	261	1	KLK8_RAT	rattus norv
36	98	4.7	873	1	FAS2_DROME	drosophila
37	98	4.7	1010	1	CONI_CHICK	gallus gall
38	97.5	4.7	1449	1	VGI2_CVPMI	porcine tra
39	97.5	4.7	1906	1	KMLS_CHICK	gallus gall
40	97	4.6	198	1	DJC5_MOUSE	mus musculu
41	97	4.6	439	1	SY62_DISOM	discopyge o
42	96.5	4.6	338	1	P24506	LAMP_CHICK	gallus gall
43	96.5	4.6	880	1	TYO3_CHICK	gallus gall
44	96.5	4.6	880	1	TYO3_RAT	mus musculu
45	96.5	4.6	1147	1	KMLS_RABIT	oryctolagus

P29294 oryctolagus

ALIGNMENTS

RESULT 1
RAGE_HUMAN

ID RAGE_HUMAN STANDARD; PRT; 404 AA.
AC Q15109; Q15279; Q9Y3R3; Q9HX7;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Advanced glycosylation end product-specific receptor precursor
DE (Receptor for advanced glycosylation end products).
GN AGER OR RAGE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Lung;
RX MEDLINE=92137587; PubMed=7835890;
RA Neeper M., Schmidt A.M., Brett J., Yan S.D., Wang F., Pan Y.C.,
RA Elliston K., Stern D., Shaw A.;
RT "Cloning and expression of a cell surface receptor for advanced
RT glycosylation end products of proteins.";
RL J. Biol. Chem. 267:14998-15004(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=95137587; PubMed=7835890;
RA Sugaya K., Fukagawa T., Matsumoto K., Mita K., Takahashi E., Ando A.,
RA Inoko H., Ikemura T.;
RT "Three genes in the human MHC class III region near the junction with
RT the class II: gene for receptor of advanced glycosylation end
RT products, PBX2 homeobox gene and a notch homolog, human counterpart
RT of mouse mammary tumor gene int-3.";
RL Genomics 23:408-419(1994).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Rowen L., Bankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA Banta A., Spies T., Hood L.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT ARG-100.
RA Abedin M.J., Yonekura H., Migita H., Karasawa J., Yamamoto Y.,
RA Yamamoto H.;
RT "Molecular heterogeneity of the receptor for advanced glycation
RT endproducts.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Malherbe P., Richards J., Gaillard H., Thompson A., Diener C.,
RA Schuler A., Huber G.;
RT "cDNA cloning of a novel secreted isoform of the human Receptor for
RT Advanced Glycation End products (RAGE) and characterization of cells
RT co-expressing cell-surface scavenger receptors and Swedish mutant
RT amyloid precursor protein.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN [6]
RP SEQUENCE FROM N.A. (ISOFORM 1).

Q14162 homo sapien
P36374 rattus norv
P34082 drosophila
P14781 gallus gall
P33470 porcine tra
P11799 gallus gall
P54101 mus musculu
P24506 discopyge o
Q98919 gallus gall
P55144 mus musculu
P55146 rattus norv
P29294 oryctolagus

```
RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 1-12 FROM N.A.
RA   Hudson B.I., Futers T.S.;
RT   "Novel polymorphisms in the receptor for advanced glycation
RT   end-products (RAGE) gene.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MEDIATES INTERACTIONS OF ADVANCED GLYCOSYLATION END
CC       PRODUCTS (AGE). THESE ARE NONENZYMATICALLY GLYCOSYLATED PROTEINS
CC       WHICH ACCUMULATE IN VASCULAR TISSUE IN AGING AND AT AN ACCELERATED
CC       RATE IN DIABETES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1).
CC       Secreted (isoform 2).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2/RAGESEC;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ENDOTHELIAL CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -----------------------------------------------------------------------
```

```
DR   EMBL; M91211; AAA03574.1; -.
DR   EMBL; D28769; BAA05958.1; -.
DR   EMBL; U89336; AAB47491.1; -.
DR   EMBL; AB036432; BAA89369.1; -.
DR   EMBL; AJ133822; CAB43108.1; -.
DR   EMBL; BC020669; AAH20669.1; -.
DR   EMBL; AF208289; AAG35728.1; -.
DR   Genew; HGNC:320; AGER.
DR   MIM; 600214; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IG2; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Immunoglobulin domain; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Alternative splicing; Polymorphism.
FT   SIGNAL        1     22
FT   CHAIN        23    404       ADVANCED GLYCOSYLATION END PRODUCT-
FT                                SPECIFIC RECEPTOR.
FT   DOMAIN       23    342       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    343    363       POTENTIAL.
FT   DOMAIN      364    404       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       31    106       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      137    215       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      252    308       IG-LIKE C2-TYPE DOMAIN 2.
FT   DISULFID     38     99       POTENTIAL.
FT   DISULFID    144    208       POTENTIAL.
FT   DISULFID    259    301       POTENTIAL.
FT   CARBOHYD     25     25       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     81     81       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DOMAIN      380    384       POLY-GLU.
FT   VARSPLIC     54     67       MISSING (IN ISOFORM 2).
FT   VARSPLIC    275    404       GVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRA
FT                                VSISIEPGEEGPTAGSVGGSGLGTLALALGILGGLLGTAAL
FT                                LIGVILWQRQRQRGEERKAPENQEEEEERAELNQSEEPEAG
FT                                ESSTGGP -> VSDLERGAGRTRRGAGNCRLCGRIRAGNSS
FT                                PGPGDPGPGDSRPAHWGHLVAKAATPRGEEGPRKPGGRG
FT                                GACRTESVGGT (IN ISOFORM 2).
FT   VARIANT     100    100       Q -> R.
FT                                /FTId=VAR_011338.
FT   CONFLICT      1      1       M -> G (IN REF. 1).
```

```
SQ   SEQUENCE   404 AA;  42802 MW;  0D584C4436C30CCE7 CRC64;

Qy   Query Match     7.0%;  Score 145.5;  DB 1;  Length 404;
     Best Local Similarity  23.5%;  Pred. No. 0.00018;
     Matches  77;  Conservative  34;  Mismatches  94;  Indels  123;  Gaps  15;

Qy    64  EPIITNDRFTS-----QRYDQGGNFT--SEMIIHNVEPSDSGNIR---CSLQNSRLHGS  112
              || |   ||    ||| |  ||   |  :| |: |  | ||    || :|   ||
Db   162  KPLVPNEKGVSVKEQTRRHPETGLFTLQSELM--VTPARGGDPRPFSCSFSPGLPRHR  218

Qy   113  AYILVQVMGELFIP-----SVNLVVAENEP----------CEVTCLPSHWTWLPDIS  154
Db   219  ALRTAPIQDPRVWEPVPLEEVQLVV--EPEGAVAPGGTVLTCEVPAQPS----PQIH  270

Qy   155  WELGLLVSHSSSYYFVPEPSDLQSAVSIIALTPQSNGTLTCVATWKSLKARKSATVNLTVI  214
Db   271  WMKD---GVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISII  321

Qy   215  RCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLT------PTCLTIRCC  267
Db   322  E-PGEEG-------------PTAGSVGGSGLGTLALALGILGGLLGTAALLIGVI  361

Qy   268  CCRRRCGCNCCCRCCPCCRRKRRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQ  325
Db   362  LWQRR---------QRGEERKAPENQEEEEERAELN-------  389

Qy   326  KTTDTASLPPKSCESSDPEQRNSSCGPP  353
Db   390  --------QSEEPEAGESSTGGP  404
```

```
RESULT 2
DCC_HUMAN
ID   DCC_HUMAN               STANDARD;      PRT;   1447 AA.
AC   P43146;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tumor suppressor protein DCC precursor (Colorectal cancer suppressor).
GN   DCC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95011532; PubMed=7926722;
RA   Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA   Vogelstein B.;
RT   "The DCC gene product in cellular differentiation and colorectal
RT   tumorigenesis.";
RL   Genes Dev. 8:1174-1183(1994).
RN   [2]
RP   SEQUENCE OF 1-750 FROM N.A.
RX   MEDLINE=90100559; PubMed=2294591;
RA   Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA   Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA   Vogelstein B.;
RT   "Identification of a chromosome 18q gene that is altered in
RT   colorectal cancers.";
RL   Science 247:49-56(1990).
RN   [3]
RP   SEQUENCE OF 107-472 FROM N.A. (SCRAMBELD EXONS).
RX   MEDLINE=91121517; PubMed=1991322;
RA   Nigro J.M., Cho K.R., Fearon E.R., Kern S.E., Ruppert J.M.,
RA   Oliner J.D., Kinzler K.W., Vogelstein B.;
RT   "Scrambled exons.";
RL   Cell 64:607-613(1991).
RN   [4]
RP   GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375.
RX   MEDLINE=94245241; PubMed=8188295;
RA   Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
RA   Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
```

```
RT   "The DCC gene: structural analysis and mutations in colorectal
RT   carcinomas.";
RL   Genomics 19:525-531(1994).
RN   [5]
RP   VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
RX   MEDLINE=94243823; PubMed=8187090;
RA   Miyake S., Nagai K., Yoshino K., Oto M., Endo M., Yuasa Y.;
RT   "Point mutations and allelic deletion of tumor suppressor gene DCC in
RT   human esophageal squamous cell carcinomas and their relation to
RT   metastasis.";
RL   Cancer Res. 54:3007-3010(1994).
CC   -!- DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
CC       DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
CC       EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
CC       POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
CC       METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
CC   -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
CC       NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X76132; CAA53735.1; -.
DR   EMBL; M32292; AAA35751.1; -.
DR   EMBL; M32286; AAA35751.1; -.
DR   EMBL; M32288; AAA52175.1; -.
DR   EMBL; M32290; AAA52176.1; -.
DR   EMBL; M63696; AAA52177.1; -.
DR   EMBL; M63700; AAA52178.1; -.
DR   EMBL; M63702; AAA52179.1; -.
DR   EMBL; M63718; AAA52180.1; -.
DR   EMBL; M63698; AAA52181.1; -.
DR   PIR; A54100; A54100.
DR   PIR; A40098; A40098.
DR   PIR; A38442; A38442.
DR   HSSP; P56276; 1IYLK.
DR   Genew; HGNC:2701; DCC.
DR   MIM; 120470; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
KW   Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW   Repeat; Anti-oncogene; Disease mutation; Polymorphism.
FT   SIGNAL        1     25       POTENTIAL.
FT   CHAIN        26   1447       TUMOR SUPPRESSOR PROTEIN DCC.
FT   DOMAIN       26   1097       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1098   1122       POTENTIAL.
FT   DOMAIN     1123   1447       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       54    124       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      154    219       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      254    317       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      345    407       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      426    522       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      525    618       FIBRONECTIN TYPE-III 2.

FT   DOMAIN      619    716       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      722    816       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      840    940       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      941   1042       FIBRONECTIN TYPE-III 6.
FT   DISULFID     61    117       BY SIMILARITY.
FT   DISULFID    161    212       BY SIMILARITY.
FT   DISULFID    261    310       BY SIMILARITY.
FT   DISULFID    352    400       BY SIMILARITY.
FT   CARBOHYD     94     94       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    628    628       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    702    702       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    168    168       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT     168    168       M -> T (IN OESOPHAGEAL CARCINOMA).
FT                               /FTId=VAR_003909.
FT   VARIANT     201    201       R -> G.
FT                               /FTId=VAR_003910.
FT   VARIANT    1375   1375       P -> H (IN A COLORECTAL CARCINOMA).
FT                               /FTId=VAR_003911.
FT   CONFLICT    138    138       MISSING (IN REF. 3).
FT   CONFLICT    233    329       MISSING (IN REF. 3).
FT   CONFLICT    421    421       MISSING (IN REF. 3).
SQ   SEQUENCE   1447 AA;  158456 MW;  4A8612766ED0471F CRC64;
```

```
Query Match       6.5%;  Score 136;  DB 1;  Length 1447;
Best Local Similarity   25.1%;  Pred. No. 0.0052;
Matches    62;  Conservative   34;  Mismatches 101;  Indels    50;  Gaps    10;

Qy    9 RDPPGSGSGNEV------------------IEGPQNARVLKGSQARFNCTVSQGW-KLIMW 50
          ::| ||  |||         :|| || :|:| | |||| | ||||
Db  215 RNPASSRTGNEAEAVRILSDPGLHRQLYFLQRPSNVVAIEGKDAVLECCVS-GYPPPSFTW 273

Qy   51 ALSDMVVLSVRPMEPIIITNDRFTSQRYDGGGNFTSEMIIHNVEPSDSDSGNIRC--SLONSR 108
          |:| | |:|| ||  :|| |||||  |                    |
Db  274 LRGEEVI------------------QLRSKKYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNEN 318

Qy  109 LHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTWLPDISW-ELGLLVSHSSYV 167
          | || ||| | | |||||  |  || |  |:|||   |||  ||  |:| ||  |: |
Db  319 ISASAELTVLPVPWFLNHPSNLYAYESMDIEFECTVSGKP-VPTVNMKNGDVVIPSDYF 377

Qy  168 FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIP 227
          |:|::  | | | |:: ||  |||||  |||||||| |||  |  :||
Db  378 QIVGGSNLR----ILGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS------ 426

Qy  228 GVYLSSLP 234
          |||  |
Db  427 -VLPSAP 432
```

```
RESULT 3
DCC_MOUSE
ID   DCC_MOUSE     STANDARD;     PRT;   1447 AA.
AC   P70211;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tumor suppressor protein DCC precursor.
GN   DCC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=96112625; PubMed=8570174;
RA   Cooper H.M., Armes P., Britto J., Gad J., Wilks A.F.;
RT   "Cloning of the mouse homologue of the deleted in colorectal cancer
RT   gene (mDCC) and its expression in the developing mouse embryo.";
RL   Oncogene 11:2243-2254(1995).
RN   [2]
RP   REVISIONS.
```

```
RC   STRAIN=BALB/c; TISSUE=Brain;
RA   Cooper H.M.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; A (shown here) and B; are
CC       produced by alternative splicing.
CC       produced by alternative initiation. A third isoform: C is
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
CC       DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
CC       BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC       ISOFORM C IS EXPRESSED ONLY IN THE EMBRYO.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS
CC       EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
CC       AND REMAIN AT THIS LEVEL IN THE ADULT.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS. DCC
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X85788; CAA59786.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:94869; Dcc.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IG2; 3.
KW   Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW   Repeat; Anti-oncogene; Alternative initiation.; Alternative splicing.
FT   SIGNAL          1     25
FT   CHAIN          26   1447       TUMOR SUPPRESSOR PROTEIN DCC.
FT   CHAIN          85   1447       TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM A.
FT   INIT_MET       85     85
FT   DOMAIN         26   1097       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     1098   1122       POTENTIAL.
FT   DOMAIN       1123   1447       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN         54    124       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN        154    219       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN        254    317       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN        345    407       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN        426    522       FIBRONECTIN TYPE-III 1.
FT   DOMAIN        525    618       FIBRONECTIN TYPE-III 2.
FT   DOMAIN        619    716       FIBRONECTIN TYPE-III 3.
FT   DOMAIN        722    816       FIBRONECTIN TYPE-III 4.
FT   DOMAIN        840    940       FIBRONECTIN TYPE-III 5.
FT   DOMAIN        941   1042       FIBRONECTIN TYPE-III 6.
FT   DISULFID       61    117       BY SIMILARITY.
FT   DISULFID      161    212       BY SIMILARITY.
FT   DISULFID      261    310       BY SIMILARITY.
FT   DISULFID      352    400       BY SIMILARITY.
FT   CARBOHYD       60     60       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD       94     94       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      299    299       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      318    318       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      478    478       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      525    522       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      628    628       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      702    702       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC      819    838       MISSING (IN ISOFORM C).
```

```
SQ   SEQUENCE   1447 AA;  158298 MW;  0D1F1097C22D5B9F CRC64;

     Query Match      Best Local Similarity   25.5%;
     Score 132;  DB 1;  Length 1447;
     Matches   56;  Conservative   31;  Mismatches   91;  Indels   42;  Gaps    9;

Qy       9  RDPGSGSGNEV-----------------IEGPQNARVLKGSQARFNCTVSQGGW--KLIMW   50
               |:|   |||                 :||||  |  ||:|:  :|| ||    ||
Db     215  RNPASIRGNEAEVRILSDPGLHRQLYFLQRPSNVIAIEGKDAVLECCVS-GYPPSFTW    273

Qy      51  ALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSR  108
                |:| ||   :       :|: |         :|  |:  || |||:|:| |   :|  |
Db     274  LRGEEVI----------------QLRSKKYSLLGG--SNLLISNVTDDDSGYTYKNEN    318

Qy     109  LHGSAYLLVQVWGELFIPSVNLVVAENEPCEVTCLPSHWTWLPDISW-ELGLLVSHSSYY  167
                |:|||   ||| ||: |   |: :||  ||:|| | :  |  |:    ||||| :|||
Db     319  ISASAELIVLVPPWFLNHPSNLYAYESMDIEFECAVSGKP-VPIVWMKNGDVVIPSDYF    377

Qy     168  FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSA  207
                :||  |  |       || | |  ||||:|:  :|  |
Db     378  QIVGGSNLR---ILGVVKSDEGFYQCVAENEAGNAQSSA  413
```

```
RESULT 4
KMLS_HUMAN
ID   KMLS_HUMAN     STANDARD;     PRT;   1914 AA.
AC   Q15746; Q14844; Q16794; O95796; O95797; O95798; O95799; Q9UBG5;
AC   Q9UTT9;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin light chain kinase, smooth muscle and non-muscle isozymes
DE   (EC 2.7.1.117) (MLCK) (Contains: Telokin (Kinase related protein)
DE   (KRP)).
GN   MYLK OR MLCK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Umbilical vein endothelial cells;
RX   MEDLINE=97304466; PubMed=9160829;
RA   Garcia J.G.N., Lazar V.L., Gilbert-Mcclain L.I., Gallagher P.J.,
RA   Verin A.D.;
RT   "Myosin light chain kinase in endothelium: molecular cloning and
RT   regulation.";
RL   Am. J. Respir. Cell Mol. Biol. 16:489-494(1997).
RN   [2]
RP   REVISIONS.
RA   Birukov K.G., Garcia J.G.N.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 2; 3A; 3B AND 4).
RC   TISSUE=Umbilical vein;
RX   MEDLINE=99216419; PubMed=10198165;
RA   Lazar V.L., Garcia J.G.N.;
RT   "A single human myosin light chain kinase gene (MLCK; MYLK).";
RL   Genomics 57:256-267(1999).
RN   [4]
RP   REVISIONS (ISOFORM 2).
RA   Birukov K.G., Garcia J.G.N.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 923-1914 FROM N.A.
RC   TISSUE=Hippocampus;
RX   MEDLINE=96121365; PubMed=8575746;
RA   Potier M.-C., Chelot E., Pekarsky Y., Gardiner K., Rossier J.,
RA   Turnell W.G.;
RT   "The human myosin light chain kinase (MLCK) from hippocampus:
RT   cloning, sequencing, expression, and localization to 3qcen-q21.";
RL   Genomics 29:562-570(1995).
RN   [6]
```

```
RP   SEQUENCE OF 1614-1914 FROM N.A.
RC   TISSUE=Lung, and Placenta;
RX   MEDLINE=20007838; PubMed=10536370;
RA   Watterson D.M., Schavocky J.P., Guo L., Weiss C., Chlenski A.,
RA   Shrinsky V.P., Van Eldik L.J., Haiech J.;
RT   "Analysis of the kinase-related protein gene found at human chromosome
RT   3q21 in a multi-gene cluster: organization, expression, alternative
RT   splicing and polymorphic marker.";
RL   J. Cell. Biochem. 75:481-491(1999).
RN   [7]
RP   SEQUENCE OF 1456-1914 FROM N.A.
RC   TISSUE=Placenta;
RA   Watterson M.D.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
CC       SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
CC       IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
CC       FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS. CENTRAL
CC       DETERMINANT IN THE DEVELOPMENT OF VASCULAR PERMEABILITY AND TISSUE
CC       EDEMA FORMATION. IN THE NERVOUS SYSTEM IT HAS BEEN SHOWN TO
CC       CONTROL THE GROWTH INITIATION OF ASTROCYTIC PROCESSES IN CULTURE
CC       AND TO PARTICIPATE IN TRANSMITTER RELEASE AT SYNAPSES FORMED
CC       BETWEEN CULTURED SYMPATHETIC GANGLION CELLS. CRITICAL PARTICIPANT
CC       IN SIGNALING SEQUENCES THAT RESULT IN FIBROBLAST APOPTOSIS.
CC   -!- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC       light-chain] phosphate.
CC   -!- SUBUNIT: TELOKIN BINDS CALMODULIN.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; A NON-MUSCLE FORM (THE
CC       LONGEST FORM), A SMOOTH-MUSCLE FORM AND TELOKIN (A C-TERMINAL
CC       SECTION WITH NO CATALYTIC ACTIVITY); ARE PRODUCED BY ALTERNATIVE
CC       INITIATION.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; 1 (SHOWN HERE), 2, 3A,
CC       3B, 4 AND DEL-1790; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES ARE
CC       EXPRESSED IN A WIDE VARIETY OF ADULT AND FETAL TISSUES AND IN
CC       CULTURED ENDOTHELIUM WITH QUALITATIVE EXPRESSION APPEARING TO BE
CC       NEITHER TISSUE- NOR DEVELOPMENT-SPECIFIC. NON-MUSCLE ISOFORM 2 IS
CC       THE DOMINANT SPLICE VARIANT EXPRESSED IN VARIOUS TISSUES. TELOKIN
CC       HAS BEEN FOUND IN A WIDE VARIETY OF ADULT AND FETAL TISSUES.
CC   -!- PTM: MLCK IS PROBABLY DOWN-REGULATED BY PHOSPHORYLATION.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U48959; AAC18423.2; -.
DR   EMBL; AF069601; AAD15921.2; -.
DR   EMBL; AF069602; AAD15921.2; -.
DR   EMBL; AF069603; AAD15922.1; -.
DR   EMBL; AF069604; AAD15924.1; -.
DR   EMBL; X85337; CAA59685.1; -.
DR   EMBL; AF096771; AAD51380.1; -.
DR   EMBL; AF096766; AAD51380.1; JOINED.
DR   EMBL; AF096767; AAD51380.1; JOINED.
DR   EMBL; AF096768; AAD51380.1; JOINED.
DR   EMBL; AF096769; AAD51380.1; JOINED.
DR   EMBL; AF096770; AAD51380.1; JOINED.
DR   EMBL; AF096774; AAD54018.1; -.
DR   EMBL; AF096771; AAD51381.1; -.
DR   EMBL; AF096771; AAD51381.1; JOINED.
DR   EMBL; AF096769; AAD51381.1; JOINED.
DR   EMBL; AF096770; AAD51381.1; JOINED.
DR   EMBL; X9870; CAA62378.1; -.
DR   HSSP; O63450; 1A06.
DR   Genew; HGNC:7590; MYLK.
DR   MIM; 600922; -.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003599; Ig_like.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   Pfam; PF00041; fn3; 3.
DR   Pfam; PF00041; ig; 8.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00047; ig; 8.
DR   PRINTS; PR00014; FNTYPEIII.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00060; FN3; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 8.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
FT   INIT_MET   1761   1761
FT   DOMAIN     1343   1413   MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE
FT                            ISOZYME.
FT   DOMAIN     1464   1719   TELOKIN.
FT   DOMAIN     1711   1774   FOR TELOKIN.
FT   DOMAIN     1824   1891   FIBRONECTIN TYPE-III.
FT   NP_BIND    1470   1478   PROTEIN KINASE.
FT   BINDING    1493   1493   CALMODULIN-BINDING.
FT   ACT_SITE   1585   1585   IG-LIKE C2-TYPE DOMAIN.
FT   DOMAIN     1906   1914   ATP (BY SIMILARITY).
FT   DOMAIN     868    998    ATP (BY SIMILARITY).
FT   DOMAIN     868    895    BY SIMILARITY.
FT   REPEAT     896    923    POLY-GLU.
FT   REPEAT     924    951    5 X 28 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT     952    979    I-1.
FT   REPEAT     980    998    I-2.
FT   DOMAIN     999    1063   I-3.
FT   DOMAIN     999    1003   I-4.
FT   REPEAT     1004   1015   I-5.
FT   REPEAT     1016   1027   6 X 12 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT     1028   1039   II-1.
FT   REPEAT     1040   1051   II-2.
FT   REPEAT     1052   1063   II-3.
FT   VARSPLIC   437    506    II-4.
FT                            II-5.
FT                            II-6.
FT   VARSPLIC   1433   1439   VSGIPKPEVAWFLEGTPVRRQEGSIEVYEDAGSHYLCLLKA
FT                            RTRDSGTYSCTASNAQGQVSCSWTLQVER -> G (IN
FT                            ISOFORM 2 AND ISOFORM 3B).
FT   VARSPLIC   1473   1546   DEVEVSD -> MKWRCQT (IN ISOFORM 3A,
FT                            ISOFORM 3B AND ISOFORM 4).
FT   CHAIN      1761   1914   GKFGQVFRLVEKKTRKVWAGKFFKAYSAKEKENIRQEISM
FT                            NCLHHPKLVQCVDAFEEKANIVMVLEIVSGGEL -> L
FT                            (IN ISOFORM 4).
FT   INIT_MET   923    923
FT   CHAIN      923    1914
FT   CHAIN      1      1914
FT   VARSPLIC   1655   1705   MISSING (IN ISOFORM 3A AND ISOFORM 3B).
FT   VARSPLIC   1790   1790   MISSING (IN ISOFORM DEL-1790).
FT   CONFLICT   933    933    V -> M (IN REF. 5).
FT   CONFLICT   963    963    S -> G (IN REF. 5). AAD15922).
FT   CONFLICT   1022   1022   P -> A (IN REF. 5).
FT   CONFLICT   1048   1050   KPM -> EAH (IN REF. 5).
FT   CONFLICT   1162   1162   P -> L (IN REF. 5). AAD15922/AAD15923).
FT   CONFLICT   1210   1210   L -> P (IN REF. 5). AAD15922/AAD15923).
FT   CONFLICT   1280   1280   E -> V (IN REF. 5). AAD15922/AAD15923).
FT   CONFLICT   1284   1284   M -> I (IN REF. 3; AAD15922/AAD15923).
FT   CONFLICT   1300   1300   A -> G (IN REF. 5).
FT   CONFLICT   1316   1316   A -> S (IN REF. 5).
FT   CONFLICT   1326   1326   T -> S (IN REF. 5).
FT   CONFLICT   1478   1478   V -> C (IN REF. 5).
```

```
FT   CONFLICT   1511   1511   S -> T (IN REF. 3; AAD15922/AAD15923).
FT   CONFLICT   1563   1563   I -> T (IN REF. 5).
FT   CONFLICT   1609   1609   A -> V (IN REF. 5).
FT   CONFLICT   1639   1639   G -> P (IN REF. 5).
FT   CONFLICT   1639   1639   G -> R (IN REF. 5).
FT   CONFLICT   1639   1640   GY -> D (IN REF. 3; AAD15922/AAD15923).
```

Query Match
Best Local Similarity   20.3%;   Score 126; DB 1; Length 1914;
Matches 86; Conservative 58; Mismatches 149; Pred. No. 0.047; Indels 130; Gaps 19;

```
Qy   24 PQNARVLKGSQARFNCTYSQGW-KLIMWALSDMVVLSVRPMEPIITNDRFTSQRVDQGG  81
Db   39 PRNLCIKEGATAKPEGRV-RGYPEPQVTWH--------------RNCQPITSGGRFL---LDCGI  85

Qy   82 NFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQV-----------MGEL 123
Db   86 RGTFSLVIHAVHEEDRGKYTCEATNG-SGARQVTVELTVEGSFAKQLGQPVVSKTLGDR 143

Qy  124 FI------PSV----------NLLVAENE-----PCEYTCLPSHWTWLPDISWELG- 158
Db  144 FSASAVETRPSIWGECPPKFAATKLGRVVVKEQMGRFSCKITIGRPQ----PQVTWLKGN 198

Qy  159 LIVSHSSYYFVPEPSDLQSAVSILALIPQSNGTLTCVATWKSLKARKSA--------TVN 210
Db  199 VPLQPSARVSVSEKNGMQ-VLEIHGVNQDDVGVYTCLVVNGSGKASMSAELSIQQLDSAN 257

Qy  211 LTVIRCPQDTGGGI----NIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPICTLIIR 265
Db  258 RSFVRETKATNSDVRKEVTNVISKESKLDSL- 288

Qy  266 CCCRRRCCGCNCCCRCCFCCRRKRGFRIQFQKKSEKEKTNKETETESGNEN----SGYNS 322
Db  289 -----EAAAKSKNCSSPQRGGSPPWAANSQPQPP-RESKLESCKDSPRTAPQTP 336

Qy  323 DEQKTTDTASLPPKSCESSDPEQRNSSCGPPHQRADQR-----PPRPASHPQASFNLASPE 378
Db  337 VLQKTSSSITLQQAARYQ----PEPRAPGLGVLSPSGGEERKRPAPPRPATPPTRQPGLGSQD 393

Qy  379 KVS 381
Db  394 VVS 396
```

RESULT 5
G55A_CHICK
```
ID   G55A_CHICK     STANDARD;      PRT;   337 AA.
AC   Q9889Z;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neurite inhibitor GP55-A precursor (OBCAM protein gamma isoform).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97157768; PubMed=9004047;
RA   Wilson D.J.A., Kim D.-S., Clarke G.A., Marshall-Clarke S., Moss D.J.;
RT   "A family of glycoproteins (GP55), which inhibit neurite outgrowth,
RT   are members of the Ig superfamily and are related to OBCAM,
RT   neurotrimin, LAMP and CEPU-1.";
RL   J. Cell Sci. 109:3129-3138(1996).
RN   [2]
RP   SEQUENCE OF 64-337 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RA   Lodge A.P., Kim D.-S., Howard M.R., McNamee C.J., Smith N., Moss D.J.;
RT   "Cloning of CEPU-s, a secreted isoform of CEPU-1, and OBCAM cDNAs from
RT   chick: structural diversity of IgLON family proteins.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INHIBITS NEURITE OUTGROWTH.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: RESTRICTED TO THE NERVOUS SYSTEM.
CC   -!- DEVELOPMENTAL STAGE: INCREASES DURING DEVELOPMENT FROM VERY LOW
CC       LEVELS AT EMBRYONIC DAY 10 AND IS MOST ABUNDANT AFTER HATCHING.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
```
```
DR   EMBL; Y08170; CAB41420.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig-c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL       1     20       BY SIMILARITY.
FT   CHAIN       21    317       NEURITE INHIBITOR GP55-A (POTENTIAL).
FT   PROPEP     318    337       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN      43    115       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN     143    201       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN     229    295       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID    50    108       POTENTIAL.
FT   DISULFID   150    194       POTENTIAL.
FT   DISULFID   236    288       POTENTIAL.
FT   CARBOHYD   133    133       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   133    133       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   277    277       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   285    285       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   298    298       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   337 AA;  36887 MW;  BAE71751856551E CRC64;
```

Query Match
Best Local Similarity   25.4%;   Score 125; DB 1; Length 337;
Matches 64; Conservative 33; Mismatches 97; Pred. No. 0.0066; Indels 337; Gaps 14;

```
Qy   26 NARVLKGSQARFNCTYSQGWKLIMWALSDMVVL-------SVRPMEPIITNDRFTSQRYD  78
Db   37 NVTVRQGESAILRCTIVDDRVRRVAW-LNRSTILYAGNDKWSIDNRVVLLSN---TKTQY  91

Qy   79 QGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRLHGSAYLTVQVMGELFIPSVNLVV 132
Db   92 --SIKIHNVDVYDEGPYTCSVQTDNHPKTSRVH----LIVQVPPQIYNISSDITV 140

Qy  133 AENEPCEVTCL-----PSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALIPQS 188
Db  141 NEGSSVTLMCLAFGRPE-----PTVTWR---HLSGKGQGFVSEDEVLE----ITGITREQ 188

Qy  189 NGTLTCVATWKSLKARKSATVNLTV-----IRCPQDTGGGINIPGVL-----SSLPSLGFS 239
Db  189 SGEYECSAV-NDVAVPDVRKVKVTVNPPYISNAKNTGASVGQKGILQCERASAVPVAEFQ 247

Qy  240 LPTWGKVGLGIA 251
Db  248 ---WFKEDPRIA 256
```

RESULT 6
CEPU_CHICK
```
ID   CEPU_CHICK     STANDARD;      PRT;   353 AA.
AC   Q90773;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
```

DE   CEPU-1 protein precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96370549; PubMed=8774445;
RA   Spaltmann F., Bruemmendorf T.;
RT   "CEPU-1, a novel immunoglobulin superfamily molecule, is expressed by
RT   developing cerebellar Purkinje cells.";
RL   J. Neurosci. 16:1770-1779(1996).
CC   -!- FUNCTION: IT MAY BE A CELLULAR ADDRESS MOLECULE SPECIFIC TO
CC   PURKINJE CELLS. IT MAY REPRESENT A RECEPTOR OR A SUBUNIT OF A
CC   RECEPTOR COMPLEX.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A MAJOR FORM AND A MINOR FORM
CC   (SHOWN HERE) ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: FOUND ON THE DENDRITES, SOMATA AND AXONS OF
CC   DEVELOPING PURKINJE CELLS. UNDETECTABLE ON OTHER NEURONS LIKE
CC   GOLGI OR GRANULE CELLS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED BY DEVELOPING CEREBELLAR PURKINJE
CC   CELLS. EXPRESSION COINCIDES WITH THE GROWTH OF THE DENDRITIC TREE,
CC   AFTER PURKINJE CELLS HAVE FINISHED THEIR MIGRATION FROM THE
CC   VENTRICULAR ZONE (FROM E15 UNTIL E21) EXPRESSED IN THE ADULT.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC   SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Z72497; CAA96578.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Repeat; Signal; Alternative splicing; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Immunoglobulin domain.
FT   SIGNAL   1   28   POTENTIAL.
FT   CHAIN   29   330   CEPU-1 PROTEIN.
FT   PROPEP   331   353   REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN   48   120   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN   148   206   IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN   234   300   IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID   55   113   POTENTIAL.
FT   DISULFID   155   199   POTENTIAL.
FT   DISULFID   241   293   POTENTIAL.
FT   CARBOHYD   42   42   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   68   68   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   150   150   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   282   282   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   290   290   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   303   303   N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID   330   330   GPI-ANCHOR (POTENTIAL).
FT   VARSPLIC   310   320   MISSING (IN MAJOR ISOFORM).
SQ   SEQUENCE   353 AA; 38736 MW; 2550C48591EBBBA6 CRC64;

Query Match   5.8%; Score 122; DB 1; Length 353;
Best Local Similarity   25.3%; Pred. No. 0.012;
Matches   62; Conservative 36; Mismatches 91; Indels 56; Gaps 13;

Qy   26 NARVLKGSQARFNCTVSQ------------GW---KLIMWALSDMVLSVRPMEPIITNDRFTSQRY 77
             ::| |:| :|:|: | || |:|:| |: ::| |:| | :| |
Db   42 NVTVRQGESATLRCSVDNRVTRVAWLNRSSILYAGNDKWCLDPRVVLLANTKTQYSIQ-- 99

Qy   78 DQGGNFTSEMIIHNVEPSDSGNIRCSLQ------------NSRLHGSAYLTVQVWGELFIPSVNLV 131
           | | :| |::|:|:: |::|:: | | |||:||:|:|:| :|| |: |:
Db   100 -------------IHDVDYDEGPYTCSVQTDNHPKTSRVH----LIYQVSPKITPETSSDIS 144

Qy   132 VAENEPCEVICLPSHWTWLFD--1SWELGLLVSHSSYPVPEPSDLQSAVSILALTFQSN 189
               | | |:| |::| |:|| | :|| | || | | |: | || |:
Db   145 INEGGNVSLTCIA---TGRPDTITWTR--HISPKAVGFISEDEYLE----ITGITREQS 194

Qy   190 GFLTCVATWKSLKARKSATVNLITV-----IRCPQDTGGGINIPGVL----SSLPSLGFSL 240
           |: | |: |:||:|| ||:| |||| | ::|| |:| ::|: |
Db   195 GEYECSAS-NDVAAPVVQRVKVTVNTPPYISDAKSTGVPVPGQKGILMCEASAVPSADFQ- 252

Qy   241 PTWGK 245
           | ||
Db   253 --WYK 255

Db   42 NVTVRQGESATLRCSVDNRVTRVAWLNRSSILYAGNDKWCLDPRVVLLANTKTQYSIQ-- 99

RESULT 7
ID   NTRI_RAT   STANDARD;   PRT;   344 AA.
AC   Q62718;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Neurotrimin precursor (GP65).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 217-229.
RC   STRAIN=Sprague-Dawley;
RX   MEDLINE=95198094; PubMed=7891157;
RA   Struyk A.F., Canoll P.D., Wolfgang M.J., Rosen C.L., D'Eustachio P.,
RA   Salzer J.L.;
RT   "Cloning of neurotrimin defines a new subfamily of differentially
RT   expressed neural cell adhesion molecules.";
RL   J. Neurosci. 15:2141-2156(1995).
CC   -!- FUNCTION: NEURAL CELL ADHESION MOLECULE.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: CENTRAL NERVOUS SYSTEM.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED AT HIGH LEVELS IN SEVERAL
CC   DEVELOPING PROJECTION SYSTEMS: IN NEURONS OF THE THALAMUS,
CC   SUBPLATE, AND LOWER CORTICAL LAMINAE IN THE FOREBRAIN AND IN THE
CC   PONTINE NUCLEUS, CEREBELLAR GRANULE CELLS, AND PURKINJE CELLS IN
CC   THE HINDBRAIN.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC   SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; U16845; AAA67445.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Repeat; Signal.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
FT   SIGNAL   1   31   POTENTIAL.
FT   CHAIN   32   321   NEUROTRIMIN.
FT   PROPEP   322   344   REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN   50   122   IG-LIKE C2-TYPE DOMAIN 1.

```
FT   DOMAIN        150    208       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN        236    302       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID       57    115       POTENTIAL.
FT   DISULFID      157    201       POTENTIAL.
FT   DISULFID      243    295       POTENTIAL.
FT   CARBOHYD       44     44       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD       70     70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      152    152       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      216    216       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      284    284       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      292    292       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      305    305       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      321    321       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID         321    321       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE      344 AA;  37998 MW;  CBB39BE53B33B224 CRC64;

Query Match             5.8%;  Score 121;  DB 1;  Length 344;
Best Local Similarity  24.1%;  Pred. No. 0.014;
Matches   59;  Conservative  37;  Mismatches  93;  Indels  56;  Gaps  13;

Qy   26  NARVLKGSQARPNCTYSQGWKLIMWALSDMVVLS------VRMEPIITNDRFTSQRYDQ   79
                 ::: ::     ::::                             ::  :::
Db   44  NVTVRQGESAILRCTIDNRVTRVAWLNRSTILYAGNDKWCLDPRVVLLSN--TQTQY--   98

Qy   80  GGNFTSEMIIHNVEPSDSGNTRCSLQ--------NSRLHGSAYLTVQVMGELFIPSVNLVVA  133
                 ::                                   ::: :::
Db   99  ------SIEIQNVDVDEGPYTCSVQTDNHPKTSRVH---LIVQVSPKIVELSSDISIN  148

Qy  134  ENEPCEVTCL------PSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSN  189
                         :::                      ::  :::
Db  149  EGNNISLTCIATGRPE----PTVTWR---HISPKAVGFVSEDEYLE----LQGITREQS  196

Qy  190  GVTLTCVATWKRSLKARKSAATVNLTVIVRCP--------QDTGGGINIPGVL----SSLPSLGFSL  240
                 ::                                  :::  :::
Db  197  GEYECSAS--NDVAAPVVRVRVNVNYPPYISEAKGTGVPVGQKGTLQCEASAVPSAEFQ-  254

Qy  241  PTWGK  245
                 :::
Db  255  --WFK  257

RESULT 8
PTK7_HUMAN

ID   PTK7_HUMAN     STANDARD;      PRT;   1070 AA.
AC   Q13308; Q13417;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase-like 7 precursor (Colon carcinoma kinase-4)
DE   (CCK-4).
GN   PTK7 OR CCK4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon carcinoma, and placenta;
RX   MEDLINE=96074849; PubMed=7478540;
RA   Mossie K., Jallal B., Alves F., Sures I., Plowman G.D., Ullrich A.;
RT   "Colon carcinoma kinase-4 defines a new subclass of the receptor
RT   tyrosine kinase family.";
RL   Oncogene 11:2179-2184(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fibroblast;
RX   MEDLINE=97037064; PubMed=8882711;
RA   Park S.-K., Lee H.-S., Lee S.-T.;
RT   "Characterization of the human full-length PTK7 cDNA encoding a
RT   receptor protein tyrosine kinase-like molecule closely related to
RT   chick KLG.";
RL   J. Biochem. 119:235-239(1996).
RN   [3]
```

```
RP   SEQUENCE FROM N.A., AND REVISION TO 834.
RA   Lee S.-T., Park S.-K., Lee H.-S., Ji A.R., Jung J.W.;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY FUNCTION AS A CELL ADHESION MOLECULE. LACKS PROBABLY
CC       THE CATALYTIC ACTIVITY OF TYROSINE KINASE. MAY BE CONNECTED TO THE
CC       PATHOPHYSIOLOGY OF COLON CARCINOMAS AND/OR MAY REPRESENT A TUMOR
CC       PROGRESSION MARKER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, LIVER, PANCREAS,
CC       KIDNEY, PLACENTA AND MELANOCYTES. WEAKLY EXPRESSED IN THYROID
CC       GLAND, OVARY, BRAIN, HEART AND SKELETAL MUSCLE. ALSO EXPRESSED IN
CC       ERYTHROLEUKEMIA CELLS. BUT NOT EXPRESSED IN COLON.
CC   -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC       RECEPTOR SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U33635; AAA87565.1; -.
DR   EMBL; U40271; AAC50484.2; -.
DR   EMBL; AF447176; AAL39062.1; JOINED.
DR   EMBL; AF447157; AAL39062.1; JOINED.
DR   EMBL; AF447158; AAL39062.1; JOINED.
DR   EMBL; AF447162; AAL39062.1; JOINED.
DR   EMBL; AF447164; AAL39062.1; JOINED.
DR   EMBL; AF447167; AAL39062.1; JOINED.
DR   EMBL; AF447170; AAL39062.1; JOINED.
DR   EMBL; AF447171; AAL39062.1; JOINED.
DR   EMBL; AF447173; AAL39062.1; JOINED.
DR   EMBL; AF447174; AAL39062.1; JOINED.
DR   EMBL; AF447175; AAL39062.1; JOINED.
DR   HSSP; P08631; 1AD5.
DR   Genew; HGNC:9618; PTK7.
DR   MIM; 601890.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig-C2.
DR   InterPro; IPR003600; Ig-like.
DR   InterPro; IPR002011; RTKinaseII.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00069; pkinase; 1.
DR   Pfam; PF00047; ig; 7.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 5.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00239; RECEPTOR_TYR_KIN_II; FALSE_NEG.
KW   Receptor; Transmembrane; Signal; Glycoprotein; Cell adhesion;
KW   Immunoglobulin domain; Repeat.
FT   SIGNAL          1     30       POTENTIAL.
FT   CHAIN          31   1070       TYROSINE-PROTEIN KINASE-LIKE 7.
FT   DOMAIN         31    704       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      705    725       POTENTIAL.
FT   DOMAIN        726   1070       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN         46    108       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN        143    208       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN        239    308       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN        336    398       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN        426    488       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN        517    577       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN        606    671       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN        796   1066       PROTEIN KINASE; INACTIVE.
FT   DISULFID       53    101       BY SIMILARITY.
FT   DISULFID      150    200       BY SIMILARITY.
```

RESULT 9

ID   CD22_MOUSE   STANDARD;   PRT;   862 AA.
AC   P35329;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   B-cell receptor CD22 precursor (Leu-14) (B-lymphocyte cell adhesion
DE   molecule) (BL-CAM) (CD22 OR LYB-8).
GN   CD22 OR LYB-8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DBA/2J and BALB/c; TISSUE=Liver;
RX   MEDLINE=93315834; PubMed=8100843;
RA   Law C.-L., Torres R.M., Sundberg H.A., Parkhouse R.M.;
RA   Brannan C.I., Copeland N.G., Jenkins N.A., Clark E.A.;
RT   "Organization of the murine Cd22 locus. Mapping to chromosome 7 and
RT   characterization of two alleles.";
RL   J. Immunol. 151:175-187(1993).
CC   -!- FUNCTION: MEDIATES B-CELL INTERACTIONS. MAY BE INVOLVED IN
CC   THE LOCALIZATION OF B-CELLS IN LYMPHOID TISSUES. BINDS SIALYLATED
CC   GLYCOPROTEINS; ONE OF WHICH IS CD45.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: CD22-ALPHA AND CD22-BETA (SHOWN
CC   HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: B-LYMPHOCYTES.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 6 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.

DR   EMBL; L16928; AAA02562.1; -.
DR   MGD; MGI:88322; Cd22.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 6.
DR   SMART; SM00410; IG2; 4.
DR   SMART; SM00408; IG; 1.
KW   Glycoprotein; Cell adhesion; Transmembrane; Signal; B-cell;
KW   Repeat; Immunoglobulin domain; Alternative splicing.
FT   SIGNAL   1   21
FT   CHAIN   22   862   B-CELL RECEPTOR CD22.
FT   DOMAIN   22   702   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   703   721   POTENTIAL.
FT   DOMAIN   722   862   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN   22   142   IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN   158   235   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN   265   331   IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN   361   418   IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN   450   506   IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN   537   593   IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN   624   681   IG-LIKE C2-TYPE DOMAIN 6.
FT   DISULFID   41   171   BY SIMILARITY.
FT   DISULFID   46   106   BY SIMILARITY.
FT   DISULFID   165   229   BY SIMILARITY.
FT   DISULFID   272   324   BY SIMILARITY.
FT   DISULFID   368   411   BY SIMILARITY.
FT   DISULFID   457   499   BY SIMILARITY.
FT   DISULFID   544   586   BY SIMILARITY.
FT   DISULFID   631   674   BY SIMILARITY.
FT   CARBOHYD   105   105   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   116   116   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   139   139   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   168   168   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   265   265   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   275   275   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   378   378   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   408   408   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   460   460   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   561   561   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   589   589   N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT   15   15   A -> V (IN BALB/C).
FT   VARIANT   19   19   Q -> R (IN BALB/C).
FT   VARIANT   76   76   K -> T (IN BALB/C).
FT   VARIANT   83   86   KAEP -> ATKTEKDPES (IN BALB/C).
FT   VARIANT   90   91   PP -> LS (IN BALB/C).
FT   VARIANT   94   94   R -> G (IN BALB/C).
FT   VARIANT   102   102   R -> G (IN BALB/C).
FT   VARIANT   173   173   S -> R (IN BALB/C).
FT   VARIANT   179   179   E -> K (IN BALB/C).
FT   VARIANT   186   186   Q -> K (IN BALB/C).
FT   VARIANT   190   190   K -> E (IN BALB/C).
FT   VARIANT   192   192   P -> S (IN BALB/C).
FT   VARIANT   236   236   P -> S (IN BALB/C).
FT   VARIANT   241   241   E -> K (IN BALB/C).
FT   VARIANT   244   244   R -> K (IN BALB/C).
FT   VARIANT   616   616   S -> I (IN BALB/C).
FT   DISULFID   246   301   BY SIMILARITY.
FT   DISULFID   343   391   BY SIMILARITY.
FT   DISULFID   433   481   BY SIMILARITY.
FT   DISULFID   524   570   BY SIMILARITY.
FT   DISULFID   613   664   BY SIMILARITY.
FT   CARBOHYD   92   92   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   116   116   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   175   175   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   184   184   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   214   214   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   268   268   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   283   283   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   405   405   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   463   463   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   567   567   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   646   646   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT   92   92   P -> R (IN REF. 2 AND 3).
FT   CONFLICT   147   147   K -> T (IN REF. 2 AND 3).
FT   CONFLICT   207   207   S -> G (IN REF. 2 AND 3).
FT   CONFLICT   495   496   VL -> RV (IN REF. 2 AND 3).
FT   CONFLICT   515   515   G -> E (IN REF. 2 AND 3).
FT   CONFLICT   881   881   E -> G (IN REF. 2 AND 3).
FT   CONFLICT   969   969   A -> P (IN REF. 2 AND 3).
FT   CONFLICT   992   992   S -> F (IN REF. 2 AND 3).
SQ   SEQUENCE   1070 AA;   118260 MW;   47CDE25B8E3698A5 CRC64;

Query Match   58.7%; Score 119; DB 1; Length 1070;
Best Local Similarity   25.3%; Pred. No. 0.084;
Matches   58; Conservative 32; Mismatches 103; Indels 36; Gaps 10;

```
Qy    20   VIEGPQNARVLKGSOARFNCTVS-QGWKLIMWALSDMVVLSVRPMEPIITNDRFTSQRYD   78
           :|:|:|        :|:| |:  : |   |  ::    ::    :| | ||||:
Db   227   VVLAPQDVVVARYEEAMFHCQFSAQPPSLQWLFEDETPITNRSRPPHLRRATVFA----  282

Qy    79   QGGNFTSEMIIHNVEPSDSGNIRCSLQNSR------LHGSAYLIVQVMGELFIPSVNLVVA  133
                     ::|:          |  |       |  | :  |  ||  |   :
Db   283   -----NGSLLLTQVRPRNAGIYRCIGQGQPGPIILEATLHLAEIEDMPLFEPRVFIAGS  337

Qy   134   ENEPCEVTCLPSHWTWLPDISWE-LGL-LVSHSSYYFVEPSDLQSAVSILALTPQSN-G  190
           | ||           |  |:|||   |   ||        |  :|     | : |||  |
Db   338   EE---RVTCLPPKGLPEPSVWWEHAGVRLPTHGRVY-----QKGHELVLANIAESDAG  387

Qy   191   TLTCVATWKSLKARKSATVNLTV------IRCPQDTGGGINIPGVLSSL  233
           || ||                       | |  |      ||   |  |
Db   388   VYTCHAA--NLAGQRQDVNITVATVPSWLKKPQDSQLEEGKPGYLDCL  434
```

```
FT   VARIANT     714    714    C -> F (IN BALB/C).
FT   VARIANT     787    787    M -> T (IN BALB/C).
FT   VARIANT     808    808    S -> T (IN BALB/C).
SQ   SEQUENCE    862 AA;  96582 MW;  A7662D6E87038E83 CRC64;
```

Query Match
Best Local Similarity 5.5%;   Score 114.5; DB 1; Length 862;
                                   Pred. No. 0.15;
Matches   86; Conservative 36; Mismatches 126; Indels 203; Gaps 18;

```
Qy   10  DPPGSGSGNEVIEGPQNARVLKGSQARFN----------CTVSQGWKL-----IMWA  51
         :| ||||                                  |            |
Db  472  NPQGSGS---------------------VLKPGVLRIQKYTWDSMPVSCAACNHKCSWALPVILNVHVA  519

Qy   52  LSDMVVLSVRPMEPIITNDRFTSQ--------------RYDQGGNFTSE----MIIHNVEPS  95
Db  520  PRDVKVLKVSPASEIRAGQRVLLQCDFAESNPAEVRFFWKKNGSLIVQEGRYLSFGSVSPE  579

Qy   96  DSGNIRCSLQNSRLHGSAYLIVQVMGELFIPSVNLV----------VAENEPC  138
Db  580  DSGNYNCMVNNS---------IGETLSQAWNLQVLYAPRRLRVSISPGDHVMEGKKA  627

Qy  139  EVTCLP-------SHWTWLPDISWELGLLVSHSSYFVPEPSDLQSAVSIIALTP---QSN  189
Db  628  TLSCEDSDANPPISQYTW----FDSSGQDLHSSGQKRLEPLEVQHT  669

Qy  190  GTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVCLG  249
Db  670  GSYRCKGT-NCIGTGESPSTLTVYSPEFIG-----------KRVALG  706

Qy  250  LAGTMLLTPTCTLTIRCCCRRCCGCNCCCRCCFCCRRKRGFRIQ------------FQK  298
Db  707  LGFCLFI--CILAI-----------WGMKIQKKWKONRSQOGLQE  738

Qy  299  KSE---------KEKTNKETETESGNENSG--------YNSDEQKTTDT  330
Db  739  NSSGQSFFVRNKKARRTPLSEGPQSQGYNPAMDDTVSIYAILRFPESDMHNAGDAGTPAT  798

Qy  331  ASLLPPKSCES----------SDPEQRNSSC  350
Db  799  QAPPPNNSDSVTYSVQIKRPMGDYENVNPSC  829
```

RESULT 10
NEO1_HUMAN

```
ID   NEO1_HUMAN     STANDARD;      PRT;   1461 AA.
AC   Q92859; O00340;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neogenin precursor.
GN   NEO1 OR NGN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97236653; PubMed=9121761;
RA   Meyerhardt J.A., Look A.T., Bigner S.H., Fearon E.R.;
RT   "Identification and characterization of neogenin, a DCC-related
RT   gene.";
RL   Oncogene 14:1129-1136(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97312699; PubMed=9169140;
RA   Vielmetter J., Chen X.-N., Miskevich F., Lane R.P., Yamakawa K.,
RA   Korenberg J.R., Dreyer W.J.;
RT   "Molecular characterization of human neogenin, a DCC-related protein,
RT   and the mapping of its gene (NEO1) to chromosomal position 15q22.3-
RT   q23.";
RL   Genomics 41:414-421(1997).
```

```
CC   -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC       TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC       DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC       MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AND ALSO IN CANCER CELL
CC       LINES.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
```

```
DR   EMBL; U61262; AAB17263.1; -.
DR   EMBL; U72391; AAC51287.1; -.
DR   HSSP; P02751; 1TTF.
     Genew; HGNC:7755; NEO1.
     MIM; 601907; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FniII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG; 1.
DR   SMART; SM00408; IGc2; 3.
KW   Transmembrane; Immunoglobulin domain; Glycoprotein; Signal; Repeat;
KW   Alternative splicing.
FT   SIGNAL        1     33       POTENTIAL.
FT   CHAIN        34   1461       NEOGENIN.
FT   DOMAIN       34   1105       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1106   1126       POTENTIAL.
FT   DOMAIN     1127   1461       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       67    136       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      166    228       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      263    327       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      355    417       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      436    533       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      536    629       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      630    729       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      735    829       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      850    950       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      951   1052       FIBRONECTIN TYPE-III 6.
FT   DOMAIN     1118   1121       POLY-VAL.
FT   DISULFID     74    129       BY SIMILARITY.
FT   DISULFID    173    221       BY SIMILARITY.
FT   DISULFID    270    320       BY SIMILARITY.
FT   DISULFID    362    410       BY SIMILARITY.
FT   CARBOHYD     73     73       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    210    210       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    326    326       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    639    639       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    715    715       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    909    909       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   1248   1300       MISSING (IN ISOFORM 2).
FT   CONFLICT    168    168       G -> N (IN REF. 2).
SQ   SEQUENCE   1461 AA;  159958 MW;  7AAE897E69635A21 CRC64;
```

Query Match   5.4%;   Score 113;   DB 1;   Length 1461;

RESULT 11

```
Best Local Similarity 23.1%;  Pred. No. 0.37;
Matches  65;  Conservative  34;  Mismatches 110;  Indels  72;  Gaps  12;

Qy   18 NEVIEGPQNARVL------KSSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITN---- 69
            |||:| :|||  |:|:| |:||: || ||:
Db  196 DRVIKLPSGMLVISNATEGDGGLYRCVVESGGGP--KYSDEVELKVLPDPEVISDLVFL 252

Qy   70 -------------------DRFTSQRYD--QGGNFTSEMIIH 90
                              :||:| |:|:  ||:|||:
Db  253 KQPSPLVRVIGQDVVLPCVASGLPTPIKWMKNEEALDTESSERLVLLAGGS----LEIS 308

Qy   91 NVEPSDSGNIRCSLQNSR--LHGSAYLTVQVMGELFIPSVNLVVAENE----PCEVTCLP 144
            |:| : | ||      |  || :|| ||:   : ||     :|    ||||| |
Db  309 DVTEDDAGTYFCIADNGNETIEAQAELTVQAQPEFLKQPTNIYAHESMDIVFECCEVTGKP 368

Qy  145 SHWTWLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILAITPQSNGTLTCVATWKSLKA 203
            :||:| |   | |  |: |  :|   || | |   |   ||:   | :||||:|
Db  369 T------PTVKWVKNGDMVIPSDYFKIVKEHNLQ----VLGLVKSDEGFYQCIAENDVGNA 419

Qy  204 RKSATVNLTVIRCPQDTGGGINIP-GVLSLSPLPSLGFSLPTW 243
            : ||  |  : | | :|| : |  |:|:| | ||:|
Db  420 QAGAQL-IILEHAPATTGPLPSAPRDVVASLVSTRFIKLIW 459
```

```
ID   OPCM_HUMAN    STANDARD;      PRT;   345 AA.
AC   Q14982;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Opioid-binding protein/cell adhesion molecule precursor (OBCAM)
DE   (Opioid-binding cell adhesion molecule) (OPCML).
GN   OPCML OR OBCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Occipital cortex;
RX   MEDLINE=95237612; PubMed=7721093;
RA   Shark K.B., Lee N.M.;
RT   "Cloning, sequencing and localization to chromosome 11 of a cDNA
RT   encoding a human opioid-binding cell adhesion molecule (OBCAM).";
RL   Gene 155:213-217(1995).
CC   -!- FUNCTION: BINDS OPIOIDS IN THE PRESENCE OF ACIDIC LIPIDS; PROBABLY
CC       INVOLVED IN CELL CONTACT.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; L34774; AAA36387.1; -.
DR   Genew; HGNC;8143; OPCML.
DR   MIM; 600632; -.
DR   InterPro; IPR003006; Ig-MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig-like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL        1     27
FT   CHAIN        28    322       OPIOID BINDING PROTEIN/CELL ADHESION
FT                                MOLECULE.
FT   PROPEP      323    345       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       50    122       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      150    209       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      237    303       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID     57    115       POTENTIAL.
FT   DISULFID    157    202       POTENTIAL.
FT   DISULFID    244    296       POTENTIAL.
FT   CARBOHYD     44     44       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     70     70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    306    306       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID       322    322       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   345 AA;  38007 MW;  E7A1I7BEA1AA3FF4 CRC64;
```

```
Query Match 65.8%;  Score 112.5;  DB 1;  Length 345;
Best Local Similarity 24.9%;  Pred. No. 0.07;
Matches  65;  Conservative  32;  Mismatches  95;  Indels  69;  Gaps  16;

Qy   26 NARVLKGSQARFNCTVSQGWKLIMWALSDMVVL------SVRPMEPIITNDRFTSQRYD 78
            |: | : |:||:|| :|  |: |:|| :|           ||||:|::| |  :||
Db   44 NVTVRQGESATLRCTIDDRVTRVAW-LNRSTLIYAGNDKWSIDPRVIILVN--TPTQY- 98

Qy   79 QGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRLHGSAYLTVQVMGELFIPSVNLVV 132
            ||||:||:|| | ||||||||||||:       |||||||||||||| |||||||:||
Db   99 -SIMIQNVDVDEGYTCSVQTDNHPKTSRVH---LIVQVPPQIMNISSDITV 147

Qy  133 AENEPCEVTCL------PSIHWTWLPDISWELGLLVSHSSYY----FVPEPSDLQSAVSIIAL 184
            | |  | :||       |: |:|:  |: :   |::|              | :| :|:|
Db  148 NEGSSVTLLCLAIGRPE----PTVTWR------HLSVKEGQQFVSEDEYLE----ISDI 192

Qy  185 TPQSNGTLTCVATWKSLKARKSATVNLIYV----IRCPQDTGGGINIPGVL---SSLPS 235
            | ||||||||||||:| |||::|||||:|      ||||||||||||||||   |||||
Db  193 KRDQSGEYECSAL-NDVAAPDVRKVKITVNYPPYISKAKNTGVSVGQKGILSCEASAVPM 251

Qy  236 LGFSLPTWGK----VGLGLAG 252
            |  : |:|       |  |:|
Db  252 AEFQ----WFKEETRLATGLDG 269
```

RESULT 12

```
ID   RAGE_BOVIN    STANDARD;      PRT;   416 AA.
AC   Q28173;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Advanced glycosylation end product-specific receptor precursor
DE   (Receptor for advanced glycosylation end products).
GN   AGER OR RAGE.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Lung;
RX   MEDLINE=92340547; PubMed=1378843;
RA   Neeper M., Schmidt A.M., Brett J., Yan S.D., Wang F., Pan Y.C.,
RA   Elliston K., Stern D., Shaw A.;
RT   "Cloning and expression of a cell surface receptor for advanced
RT   glycosylation end products of proteins.";
RL   J. Biol. Chem. 267:14998-15004(1992).
CC   -!- FUNCTION: MEDIATES INTERACTIONS OF ADVANCED GLYCOSYLATION END
CC       PRODUCTS (AGE). THESE ARE NONENZYMATICALLY GLYCOSYLATED PROTEINS
CC       WHICH ACCUMULATE IN VASCULAR TISSUE IN AGING AND AT AN ACCELERATED
CC       RATE IN DIABETES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
```

```
CC   -!- TISSUE SPECIFICITY: ENDOTHELIAL CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC
```

```
DR   EMBL; M91212; AAA03575.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
KW   Immunoglobulin domain; Glycoprotein; Transmembrane; Repeat; Signal.
FT   SIGNAL        1     22
FT   CHAIN        23    416
FT   DOMAIN       23    352      ADVANCED GLYCOSYLATION END PRODUCT-
FT                               SPECIFIC RECEPTOR.
FT   TRANSMEM    353    373      POTENTIAL.
FT   DOMAIN      374    416      CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       31    105      IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      136    214      IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      262    318      IG-LIKE C2-TYPE DOMAIN 2.
FT   DISULFID     38     98      POTENTIAL.
FT   DISULFID    143    207      POTENTIAL.
FT   DISULFID    269    311      POTENTIAL.
FT   CARBOHYD     25     25      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     80     80      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   DOMAIN      391    396      POLY-GLU.
SQ   SEQUENCE    416 AA;  44182 MW;  B703815573E767AE CRC64;

Query Match             5.4%;  Score 112.5;  DB 1;  Length 416;
Best Local Similarity  23.3%;  Pred. No. 0.088;
Matches   74;  Conservative   35;  Mismatches   95;  Indels  113;  Gaps  17;

Qy    74 SQRYDQGGNFT--SEMIIHNVEPSDSGNIR-----CS-----LONSRLHGSAYLTVQVMGE- 122
Db   176 TKRHPKTGLFLIHSELM--VTPARGGALHPTFSCSFFTPGLPRRALHTAPIQLRVWSEH 232

Qy   123 -----LEIPSVNLVVAENP---------CEVTCLPSHWTWLPDISWELG 158
Db   233 RGGEGPNVDAVPLKEVQLVV----EPEGAVAPGGTVTLICERAPQPP---PQIHW--- 281

Qy   159 LLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVTRCPQ 218
Db   282 -IKDGRPLPLPPGPMLLLPEVG-----PEDQGTYSCVATHPSHGPQESRAVSVTIIETGE 335

Qy   219 D--TGGGININPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLIPTCTLTIRCCCCRRCCGC 276
Db   336 EGTTAGSVEGPG----LETLALTLGILG--GLG-------TVALLIGVIVWHRR---- 376

Qy   277 NCCCRCCFCCRRKRGFRIQFQKKSEKKETKNKETETESGNENSGYNSDEQKTTDTASLPPK 336
Db   377 ----RQRKG------------------QERKVPENQEEEEERAELNQPE---------- 404

Qy   337 SCESSDPEQRNSSCGPP 353
Db   405 ----EPEAAESSTGGP 416
```

```
RESULT 13
ID   OPCM_BOVIN     STANDARD;      PRT;   345 AA.
AC   P11834;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Opioid-binding protein/cell adhesion molecule precursor (OBCAM)
DE   (Opioid-binding cell adhesion molecule) (OPCML).
GN   OPCML OR OBCAM OR OCAM.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=89251576; PubMed=2721489;
RA   Schofield P.R., McFarland K.C., Hayflick J.S., Wilcox J.N.,
RA   Cho T.M., Roy S., Lee N.M., Loh H.H., Seeburg P.H.;
RT   "Molecular characterization of a new immunoglobulin superfamily
RT   protein with potential roles in opioid binding and cell contact.";
RL   EMBO J. 8:489-495(1989).
CC   -!- FUNCTION: BINDS OPIOIDS IN THE PRESENCE OF ACIDIC LIPIDS; PROBABLY
CC       INVOLVED IN CELL CONTACT.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC       similarity).
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC
```

```
DR   EMBL; X12672; CAA31192.1; -.
DR   PIR; S03199; S03199.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL        1     27
FT   CHAIN        28    322      OPIOID BINDING PROTEIN/CELL ADHESION
FT                               MOLECULE.
FT   PROPEP      323    345      REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       50    122      IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      150    209      IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      237    303      IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID     57    115      POTENTIAL.
FT   DISULFID    157    202      POTENTIAL.
FT   DISULFID    244    296      POTENTIAL.
FT   CARBOHYD     44     44      N-LINKED (GLCNAC...)  (PROBABLE).
FT   CARBOHYD     70     70      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    140    140      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    285    285      N-LINKED (GLCNAC...)  (PROBABLE).
FT   CARBOHYD    293    293      N-LINKED (GLCNAC...)  (PROBABLE).
FT   CARBOHYD    306    306      N-LINKED (GLCNAC...)  (POTENTIAL).
FT   LIPID       322    322      GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE    345 AA;  37914 MW;  D1ECC8D9E7D8CB19 CRC64;

Query Match             5.3%;  Score 111.5;  DB 1;  Length 345;
Best Local Similarity  25.0%;  Pred. No. 0.084;
Matches   64;  Conservative   31;  Mismatches   96;  Indels   65;  Gaps  15;

Qy    26 NARVLKGSQARFNCTVSQGWKLIMWALSDMVVL------SVRMEPIITNDRFTSQRYD 78
Db    44 NVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAGNDKWSIDPRVIILVN--TPTQY 98

Qy    79 QGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRLHGSAYLITVQVMGELFIPSVNLVV 132
```

RESULT 14

```
ID    NEO1_CHICK     STANDARD;     PRT;    1443 AA.
AC    Q90610;
DT    16-OCT-2001 (Rel. 40, Created)
DT    16-OCT-2001 (Rel. 40, Last sequence update)
DT    15-JUN-2002 (Rel. 41, Last annotation update)
DE    Neogenin (Fragment).
OS    Gallus gallus (Chicken).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC    Gallus.
OX    NCBI_TaxID=9031;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=White leghorn; TISSUE=Embryonic brain;
RX    MEDLINE=95105243; PubMed=7806578;
RA    Vielmetter J., Roman J.M., Dreyer W.J.;
RT    "Neogenin, an avian cell surface protein expressed during terminal
RT    neuronal differentiation, is closely related to the human tumor
RT    suppressor molecule deleted in colorectal cancer.";
RL    J. Cell Biol. 127:2009-2020(1994).
CC   -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC       TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC       DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC       MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- DEVELOPMENTAL STAGE: IN RETINA, EXPRESSED ON GANGLION CELL FIBERS
CC       AS SOON AS THEY BEGIN TO EXTEND THEIR AXONS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR    EMBL; U07644; AAC59662.1; -.
DR    HSSP; P11276; 2MFN.
DR    InterPro; IPR003961; FN_III.
DR    InterPro; IPR003962; FnIII_repeat.
DR    InterPro; IPR003006; Ig_MHC.
DR    InterPro; IPR003598; Ig-C2.
DR    InterPro; IPR003600; Ig_like.
DR    Pfam; PF00041; ig; 4.
DR    Pfam; PF00047; ig; 4.
DR    PRINTS; PR00014; FNIYPEIII.
DR    SMART; SM00060; FN3; 6.
DR    SMART; SM00410; IG_like; 2.
DR    SMART; SM00408; IGc2; 2.
KW    Transmembrane; Immunoglobulin domain; Glycoprotein; Repeat.
FT    NON_TER        1      1
```

```
FT    DOMAIN        <1   1090       EXTRACELLULAR (POTENTIAL).
FT    TRANSMEM    1091   1111       POTENTIAL.
FT    DOMAIN      1112   1443       CYTOPLASMIC (POTENTIAL).
FT    DOMAIN        33    102       IG-LIKE C2-TYPE DOMAIN 1.
FT    DOMAIN       132    194       IG-LIKE C2-TYPE DOMAIN 2.
FT    DOMAIN       229    293       IG-LIKE C2-TYPE DOMAIN 3.
FT    DOMAIN       321    383       IG-LIKE C2-TYPE DOMAIN 4.
FT    DOMAIN       422    519       FIBRONECTIN TYPE-III 1.
FT    DOMAIN       522    615       FIBRONECTIN TYPE-III 2.
FT    DOMAIN       616    714       FIBRONECTIN TYPE-III 3.
FT    DOMAIN       720    814       FIBRONECTIN TYPE-III 4.
FT    DOMAIN       835    935       FIBRONECTIN TYPE-III 5.
FT    DOMAIN       936   1037       FIBRONECTIN TYPE-III 6.
FT    DISULFID      40     95       BY SIMILARITY.
FT    DISULFID     139    187       BY SIMILARITY.
FT    DISULFID     236    286       BY SIMILARITY.
FT    DISULFID     328    376       BY SIMILARITY.
FT    CARBOHYD      39     39       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     176    176       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     292    292       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     456    456       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     475    475       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     625    625       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     700    700       N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD     894    894       N-LINKED (GLCNAC...) (POTENTIAL).
SQ    SEQUENCE   1443 AA;  158050 MW;  55EC67955799C0E26 CRC64;
```

```
Query Match         5.3%;
Best Local Similarity   25.7%;
Matches   48;  Conservative   28;
```
Score 111.5; DB 1; Length 1443;
Pred. No. 0.49;
Mismatches   74; Indels   37; Gaps   11;

```
Qy   21  IEGPQNARVLKGSQARPENCTVSQGW--KLIMWALSDMVVLSVRPMEPIITNDRFTSQRY  77
             : ||  || ||: :|: :| ::|      |||  :   : |||:      :    |:
Db  218  VRQPSSLTKVTGQNAVFPC-VAGGFPIPYVRWTKNG---------EELIFED--SERFA  264

Qy   78  -DQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSAYLIVQVWGELFIPSVNLVVAE  134
          |  |||:  | |  | |||:|| ||    |||:||| |:: ||: |||  :  || | :
Db  265  LRAGGS---LIISDVTEEDVGTYTCIADNENETIEAQAELAVQVPPEFLKRPANIYAHE  320

Qy  135  NE---PCEVTCLPSHWTWLPDISW-ELGILLVSHSSYFVPEPSDLQSAVSILALTPQSN  189
             :    |    :||| ||| |  |  :    |: :| || || :|| : |  :  |:
Db  321  SMDIVFECEVTGKPT-------PTVKWVKNGDVVIPSDYFKIVEEHNLQ---VLGLVKSDE  371
```

```
Qy  190  GFYQCIA  196
          ||:|  
Db  372  GFYQCIA  378
```

RESULT 15

```
ID    SMP_COTJA     STANDARD;     PRT;    620 AA.
AC    Q91254;
DT    15-JUL-1998 (Rel. 36, Created)
DT    15-JUL-1998 (Rel. 36, Last sequence update)
DT    15-JUN-2002 (Rel. 41, Last annotation update)
DE    Schwann cell myelin protein precursor.
GN    SMP.
OS    Coturnix coturnix japonica (Japanese quail).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC    Coturnix.
OX    NCBI_TaxID=93934;
RN    [1]
RP    SEQUENCE FROM N.A., AND SEQUENCE OF 18-28; 120-132; 135-157 AND
RP    563-571.
RX    MEDLINE=92153423; PubMed=1739462;
RA    Dulac C., Tropak M.B., Cameron-Curry P., Rossier J., Marshak D.R.,
RA    Roder J., le Douarin N.M.;
RT    "Molecular characterization of the Schwann cell myelin protein, SMP:
RT    structural similarities within the immunoglobulin superfamily.";
RL    Neuron 8:323-334(1992).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EXCLUSIVELY EXPRESSED BY MYELINATING AND
```

```
Qy   99  ------SIMIQNVDVYDEGPYTCSVQVTDNHPKTSRVH----LIVQVPPQIMNISSDVTV  147
                : ||: ||  | |   |   :|  |  |     |||  |   :|:|: :|| :| 
Db  133  AENEPCEVTCL----PSHWTWLPDISWELGLLVSHSSYY----FVPEPSDLQSAVSILAL  184

Qy  148  NEGSSVTLLCLAIGRPE----PTVTWR----HLSVKEGQGFVSEDEYIE----ISDI  192
          : ||  :| |   || |     |  | |     | |: ||  ||  | |      |||
Db  185  TPQSNGTLTCVATWKSLKARKSATVNLTV-----IRCPQDTGGGINIPGVL----SSLPS  235

Qy  236  LGFSLPTWGKVGLGLA  251
         |   || ||  | | |
Db  193  KRDQSGEYECSAL-NDVAAPDVRKVKITVNPYYISKAKNTGVSVGQKGILSCEASAVPM  251

Db  252  AEFQ---WFKEDTRLA  264
```

Wed Apr 30 10:00:34 2003

```
CC   -!-      NONMYELINATING SCHWANN CELLS AND OLIGODENDROCYTES.
CC   -!-      DEVELOPMENTAL STAGE: FIRST SYNTHESIZED AT EMBRYONIC DAY 5, IT
CC            REMAINS EXPRESSED BY CULTURED SCHWANN CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
```

```
DR   EMBL; S83711; AAB21466.1; -.
DR   HSSP; P56276; 1TLK.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Myelin; Glycoprotein; Cell adhesion; Transmembrane; Signal; Repeat;
KW   Immunoglobulin domain.
FT   SIGNAL        1     17
FT   CHAIN        18    620       SCHWANN CELL MYELIN PROTEIN.
FT   DOMAIN       18    516       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    517    536       POTENTIAL.
FT   DOMAIN      537    620       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       28    106       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      151    223       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      253    311       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      339    398       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      424    495       IG-LIKE C2-TYPE DOMAIN 4.
FT   DISULFID     35    164       BY SIMILARITY.
FT   DISULFID     40     99       BY SIMILARITY.
FT   DISULFID    158    216       BY SIMILARITY.
FT   DISULFID    260    304       BY SIMILARITY.
FT   DISULFID    346    391       BY SIMILARITY.
FT   DISULFID    420    429       BY SIMILARITY.
FT   DISULFID    431    488       BY SIMILARITY.
FT   CARBOHYD    222    222       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    314    314       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    331    331       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    405    405       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    449    449       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   620 AA;  66943 MW;  004B3EC7EDC18FBA CRC64;
```

```
Query Match             5.3%;
Best Local Similarity  21.5%;                  Score  111;  DB  1;  Length  620;
  Matches   54;  Conservative   33;  Mismatches   84;  Indels   80;  Gaps   12;
                                              Pred. No. 0.19;

Qy    23  GPQNARVLKGSQARFNCTV-SQGWKLIMWALSDMVLSVRPMEPIITNDRFTSQRYDQGG   81
          ||:||      ||  |||||          |:|   :  |   |||:|
Db   246  GP--TEVEGSDVELGCEAEGRPAPLISWFRGSEVL------------------REEPGR  285

Qy    82  NFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVWGELFIPSVN--LVVAENEPCE  139
          :::||||  |             ||   |||  |    :|| |     |  |
Db   286  NL--RLLSNVGPDDGGSFSCVAENRHGRHNRSLQLRVAYAPRAPVINGSLWVVSGDPVS  343

Qy   140  VTCLPSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAV---SILALT--------  185
          |:|  |         : :  | |  | |  |     |    |    |
Db   344  VTCRAE--------------------SEPAILTVLRGGKVMAAIYEDHVTEMR  379

Qy   186  ---PQSNGTLTCVATWKSLKARKSATVNLTV--------IRCPQDTGGGINIPGVLSSL  233
          :   | |:||| ||       :  ::|          | ||  | | | |||:||
Db   380  PARPEDGGTYSCVA--ENQHGASSTSFNISVEYPPLVLPASRCTAG-CDGVRCVCMVNSI  436

Qy   234  P--SLGFSLPT  242
          |   || |||||
Db   437  PDSSLVFELPT  447
```

OM protein - protein search, using sw model

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 18:09:54 ; Search time 19.0659 Seconds
                 (without alignments)
                 1946.300 Million cell updates/sec

Title: US-09-729-264-4
Perfect score: 2088
Sequence: 1 MVAGAMENRDPPGSGSGNEV..........HPQASFNLASPEKVSNTTVV 386

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   PIR_73:*
            1:   pir1:*
            2:   pir2:*
            3:   pir3:*
            4:   pir4:*

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	145.5	7.0	404	1	I61596	advanced glycosyla
2	136	6.5	1447	2	A54100	tumor suppressor p
3	134.5	6.4	1427	1	151669	tumor suppressor -
4	133.5	6.4	871	1	I48696	protein-tyrosine k
5	133.5	6.4	881	1	I48697	protein-tyrosine k
6	129.5	6.2	164	2	T24272	hypothetical prote
7	122.5	5.9	188	2	T15651	hypothetical prote
8	121	5.8	344	2	T15551	neurotrimin - rat
9	119.5	5.7	6642	2	T29757	protein UNC-89 - C
10	119	5.7	1070	2	JC4593	protein-tyrosine k
11	115.5	5.5	946	1	A47299	ror-related recept
12	114.5	5.5	152	2	T18975	hypothetical prote
13	114.5	5.5	531	2	S20900	titin - mouse (fra
14	114.5	5.5	862	1	I19583	differentiation an
15	114.5	5.5	26926	1	I38344	titin, cardiac mus
16	112.5	5.4	345	2	JC4025	opioid-binding cel
17	112.5	5.4	416	1	A42879	advanced glycosyla
18	112	5.4	868	2	A46512	CD22 homolog/B lym
19	111.5	5.3	345	2	S03199	opioid-binding pro
20	111.5	5.3	1443	2	I50600	neogenin - chicken
21	111.5	5.3	6805	2	S20901	titin - rabbit (fr
22	111	5.3	620	2	JH0593	Schwann cell myeli
23	111	5.3	693	2	S49928	sodium-dependent p
24	110.5	5.3	338	2	JC1238	opioid-binding pro
25	110.5	5.3	345	2	JC1239	opioid-binding pro
26	110.5	5.3	4391	2	A38096	perlecan precursor
27	109.5	5.2	364	2	A30521	myeloid cell surfa
28	109	5.2	3375	2	T19821	hypothetical prote
29	107	5.1	391	2	T09058	butyrophilin homol
30	106.5	5.1	487	2	S65133	butyrophilin - mou
31	106.5	5.1	802	2	T13149	mitogen-and stress
32	106.5	5.1	841	2	CC5894	killer cell inhibi
33	106.5	5.1	1272	2	S26180	neurofascin - chic
34	106	5.1	1092	1	JN0635	neural cell adhesi
35	105.5	5.1	662	2	T16525	hypothetical prote
36	104.5	5.0	1177	2	T16594	irregular chiasm C
37	104	5.0	764	2	A49448	protein unc-52 [im
38	104	5.0	2295	2	C88369	polyprotein - fava
39	104	5.0	5825	1	T12217	polyprotein -
40	103.5	5.0	2491	1	A28372	insulin-like growt
41	103.5	5.0	3707	2	S18252	heparan sulfate pr
42	102.5	4.9	890	1	A53743	protein-tyrosine k
43	102	4.9	721	2	T41530	hypothetical prote
44	101.5	4.9	215	2	T16542	hypothetical prote
45	101.5	4.9	1449	2	S47423	E2 glycoprotein pr

ALIGNMENTS

RESULT 1
I61596
advanced glycosylation end-products receptor precursor - human
N;Alternate names: advanced glycosylation end products receptor and product-binding protein, 35K; glycoprot
C;Species: Homo sapiens (man)
C;Date: 24-May-1996 #sequence_revision 07-Feb-1997 #text_change 16-Jul-1999
C;Accession: I61596; B42879; S27968
R;Sugaya, K.; Fukagawa, T.; Matsumoto, K.; Mita, K.; Takahashi, E.; Ando, A.; Inoko,
Genomics 23, 408-419, 1994
A;Title: Three genes in the human MHC class III region near the junction with the cla
nterpart of mouse mammary tumor gene int-3.
A;Reference number: A55562; MUID:95137587; PMID:7835890
A;Accession: I61596
A;Molecule type: DNA
A;Residues: 1-404 <RES>
A;Cross-references: GB:D28769; NID:g561657; PIDN:BAA05958.1; PID:g561659
A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/
R;Neeper, M.; Schmidt, A.M.; Brett, J.; Yan, S.D.; Wang, F.; Pan, Y.C.; Elliston, K.;
J. Biol. Chem. 267, 14998-15004, 1992
A;Title: Cloning and expression of a cell surface receptor for advanced glycosylation
A;Reference number: A42879; MUID:92340547; PMID:1378843
A;Accession: B42879
A;Molecule type: mRNA
A;Residues: 'G',2-99,'R',101-404 <NEE>
A;Cross-references: EMBL:M91211; NID:g190845; PIDN:AAA03574.1; PID:g190846
A;Experimental source: lung
C;Comment: Advanced glycosylation end products are heterogeneous nonenzymatically gly
cellular function, thus contributing to tissue lesions in diabetes.
C;Comment: This receptor appears also to mediate the effects of amyloid beta peptide
ates in the neurotoxic pathway that produces dementia in Alzheimer's disease.
C;Genetics:
A;Gene: GDB:AGER
A;Cross-references: GDB:306354; OMIM:600214
A;Map position: 6p21.3-6p21.3
A;Introns: 18/1; 53/3; 119/1; 140/3; 170/1; 231/1; 274/3; 322/1; 331/1; 373/2
C;Function:
C;Description: neuronal receptor for amphoterin, a DNA-binding protein involved in ne
C;Description: advanced glycosylation end products receptor; immunoglobulin homology
C;Superfamily: advanced glycosylation end products receptor; immunoglobulin homology
C;Keywords: Alzheimer's disease; glycoprotein; receptor; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-404/Product: advanced glycosylation end products receptor #status predicted <MAT
F;23-344/Domain: extracellular #status predicted <EXT>
F;31-101/Domain: immunoglobulin homology <IM1>
F;137-210/Domain: immunoglobulin homology <IM2>
F;252-303/Domain: immunoglobulin homology <IM3>
F;345-362/Domain: transmembrane #status predicted <TMM>
F;363-404/Domain: intracellular #status predicted <INT>
F;25,81/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;38-99,144-208,259-301/Disulfide bonds: #status predicted

Query Match   7.0%;   Score 145.5;   DB 1;   Length 404;

```
Best Local Similarity    23.5%;  Pred. No. 0.0026;
Matches   77;  Conservative   34;  Mismatches   94;  Indels  123;  Gaps  15;

Qy     64 EPIITNDRFTS------------QRYDQGGNFT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGS 112
          :|:|:|:|||            ||:|:|| ||||||| |||||||||||||:|:  ||||||||
Db    162 KPLVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGDPRPTFSCSFSPGLPRHR 218

Qy    113 AVLTVQVMGELFIP------------SVNLVVAENEP------------CEVTCLPSHWTWLPDIS 154
           | |   | ||:|              ||| | |:|             |||||||||||||||:
Db    219 ALRTAPIQPRWEPVPLEEVQLVV---EPEGGAVAPGGTVVLTCEVPAQPS----PQIH 270

Qy    155 WELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVI 214
           ||||||||| |:| ||
Db    271 WMKD----------GVPLPPPSPVLLPEIGPQDQGTYSCVATHSSHGPQESRAVSISII 321

Qy    215 RCPQDTGGGINIPGVLSLPSLGFSLPTWGKVGLGLAGTMLLT------PTCTLTIRCC 267
          |||||||||||||    :                              |||||: ||
Db    322 E-PGEEG-------------------PTAGSVGGSGLGTLALAIGILGGLGTAALLIGVI 361

Qy    268 CCRRRCCGCNCCCFCCFRRKRGFRIQFQKKSEKEKT--NKETETFSGNENSGYNSDEQ 325
           | |   | |
Db    362 LWQRR--------------QRRGEERKAPENQEEEERAELN------- 389

Qy    326 KTTDTASLPPKSCESSDPEQRNSSCGPP 353
          :| | |
Db    390 ------------QSEEPEAGESSIGGP 404
```

RESULT 2
A54100
tumor suppressor protein DCC precursor - human
N;Alternate names: colorectal cancer suppressor DCC
C;Species: Homo sapiens (man)
C;Date: 02-Aug-1994 #sequence_revision 02-Aug-1994 #text_change 05-Nov-1999
C;Accession: A54100; A40098
R;Hedrick, L.; Cho, K.R.; Fearon, E.R.; Wu, T.C.; Kinzler, K.W.; Vogelstein, B.
Genes Dev. 8, 1174-1183, 1994
A;Title: The DCC gene product in cellular differentiation and colorectal tumorigenesis.
A;Reference number: A54100; MUID:95011532; PMID:7926722
A;Accession: A54100
A;Molecule type: mRNA
A;Residues: 1-1447 <HED>
A;Cross-references: EMBL:X76132; NID:g453209; PIDN:CAA53735.1; PID:g453210
Science 247, 49-56, 1990
A;Title: Identification of a chromosome 18q gene that is altered in colorectal cancers.
R;Fearon, E.R.; Cho, K.R.; Nigro, J.M.; Kern, S.E.; Simons, J.W.; Ruppert, J.M.; Hamilto
A;Reference number: A40098; MUID:90100559; PMID:2294591
A;Accession: A40098
A;Molecule type: mRNA
A;Residues: 1-750 <FEA>
A;Cross-references: GB:M32292; NID:g181492; PID:g181493
C;Genetics:
A;Gene: GDB:DCC
C;Cross-references: GDB:119838; OMIM:120470
A;Map position: 18q21.1-18q21.1
C;Keywords: transmembrane protein; tumor suppressor
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1447/Product: tumor suppressor protein DCC #status predicted <MAT>

```
Query Match    62%;  Score 136;  DB 2;  Length 1447;
Best Local Similarity    25.1%;  Pred. No. 0.051;
Matches   62;  Conservative   34;  Mismatches  101;  Indels  50;  Gaps  10;

Qy      9 RDPPGSSGGNEV------------------------IEGPQNARVLKGSQARFNCTVSQGNEV--KLIMW 50
           || |||| ||                          ||| | :||
Db    215 RNPASSRIGNEAEVRILSDPGLHRQLFVLFQRPSNVVAIEGKDAVLECCVS-GYPPSPFTW 273

Qy     51 ALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSR 108
Db    274 LRGEEVI----------------QLRSKKYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNEN 318

Qy    109 LHGSAYLTVQVGMGELFIPSVNLVVAANEPCEVTCLPSHWTWLPDISW-ELGLLVSHSSYY 167
Db    ...
```

RESULT 3
I51669
tumor suppressor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I51669
R;Pierceall, W.E.; Reale, M.A.; Candia, A.F.; Wright, C.V.; Cho, K.R.; Fearon, E.R.
Dev. Biol. 166, 654-665, 1994
A;Title: Expression of a homologue of the deleted in colorectal cancer (DCC) gene in
A;Reference number: I51669; MUID:95113183; PMID:7813784
A;Accession: I51669
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1427 <PIE>
A;Cross-references: EMBL:U10986; NID:g606873; PIDN:AAA70168.1; PID:g606874
C;Genetics:
A;Gene: XDCCa

```
Query Match     6.8%;  Score 134.5;  DB 2;  Length 1427;
Best Local Similarity    25.4%;  Pred. No. 0.065;
Matches   65;  Conservative   33;  Mismatches   93;  Indels  65;  Gaps  15;

Qy     12 PGSGG-GNEV------------------------IEGPQNARVLKGSQARFNCTVSQGW--KLIMWAL 52
           |||| |||                          ||| | :||
Db    217 PGSARVGNEAELRILSEGLHRQQVFLQRPSNVVAIEGQDAVLECAVS-GYPPTVWMQ 275

Qy     53 SDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLH 110
Db    276 GD--------EPVPIRTR---KYSVLGG--SNLLISNVTDDDAGAYTCVVATYKNENTS 320

Qy    166 YFVPEPSDLQSAVSILALTPQSNGTLTCVA------------TWKSLKARKSATVNLTVI-RC 216
Db    321 FSADLTVMYPPQFLNHPANLYAYESMDIEFECAVSGKPS------PTVKWTKNGEVVIPSD 375

Qy    217 PQDTGGGINIPGVLSS 232
Db    376 YFOIVDGSNLR---ILGLVKSDEGYYQCIAENEAGNIQTYAQLIIPDPAVPSSSILPSA 431

Qy    319 ISASAELTLVLPPWFLNHPSNLYAYESMDIEFECTYSGKP-VPTVNWMKNGDVVIPSDYF 377
Qy    168 FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTIVIRCPQDTGGGINIP 227
Db    378 QIVGGSNLR----ILGVWKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS----- 426
Qy    228 GVLSSLP 234
Db    427 -VLPSAP 432

Qy    432 PRDV------VPVLVSS 442
```

RESULT 4
I48696
protein-tyrosine kinase (EC 2.7.1.112) nsk2 precursor, splice form 2 - mouse
N;Alternate names: receptor-type tyrosine kinase
N;Contains: protein-tyrosine kinase (house mouse)
C;Species: Mus musculus (house mouse)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C;Accession: I48696
R;Ganju, P.; Walls, E.; Brennan, J.; Reith, A.D.
Oncogene 11, 281-290, 1995
A;Title: Cloning and developmental expression of Nsk2, a novel receptor tyrosine kina
A;Reference number: I48696; MUID:95349951; PMID:7624144
A;Accession: I48696
A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-871 <GAN1>
A;Cross-references: EMBL:X86444; NID:g929723; PIDN:CAA60165.1; PID:g929724
A;Experimental source: splice form 2
A;Accession: S60738

Left column

A;Molecule type: DNA
A;Residues: 1-456, 'A',466-871 <GAN2>
A;Cross-references: EMBL:X86444; NID:g929723
A;Gene: nsk2
A;Genetics:
A;Cross-references: MGI:103308
C;Genetics:
C;Superfamily: mouse ror-related receptor; immunoglobulin homology; protein kinase homol
C;Keywords: ATP; glycoprotein; phosphotransferase; receptor; transmembrane protein; tyr
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-871/Product: protein-tyrosine kinase nsk2, splice form 4 #status predic
F;22-456,'A',466-871/Product: protein-tyrosine kinase nsk2, splice form 2 #status predicted <MAT2>
F;42-101/Domain: immunoglobulin homology <IMM1>
F;135-192/Domain: immunoglobulin homology <IMM2>
F;226-284/Domain: immunoglobulin homology <IMM3>
F;498-518/Domain: protein-tyrosine kinase homology predicted <TRM>
F;575-865/Domain: protein kinase homology <KIN>
F;583-591/Region: protein kinase ATP-binding motif
F;222,462/Binding site: carbohydrate (Asn) (covalent) #status predicted

```
Query Match        6.48;     Score 133.5;   DB 1;   Length 871;
Best Local Similarity   21.6%;   Pred. No. 0.045;
Matches  73;   Conservative  45;   Mismatches  127;   Indels  93;   Gaps  17;

Qy    4 GAMENRDPPGSGSGNEVIEGPQONARVLKGSQARFNCTVSOGWK-LIMWALSDMVVLSVRP 62
          ||| ||          |:| |   | ||  |::|| :| |  |   ||:|| || |
Db  113 GALQVKMKP------KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD--------- 158

Qy   63 MEPIITNDRFTSORYDQOGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-------- 117
                                                                   :| |
Db  159 -NALRENSRIAALE-------SGSLRIHNVQKEDAGOYRCVAKNSL---GTAYSKLVKLEV 208

Qy  118 QVMGELFIPSVNLVVAAENEPCEVTC----LPSHWTWLPDISW-ELGLLVSHSSYFVPEP 172
            |    : | |  | |    |  | |       | |  |      ||||||||||||||
Db  209 EVLGRILRAPESHNVTFGSFVTLRCTEIGIP----VPTISWIENGNAVSSGSIQESVKD 263

Qy  173 SDLQSAVSILALTPQSNGTLTCVAT-----WKSLKARKSATVNLTVIR---------CPQD 219
                                         : | ||:||:| ||| ||
Db  264 RVIDSRLQLFITKP--GLYTCIATNKHGEKFSTAKAAATVSIAEWSKSQKDSQGYCAQY 320

Qy  220 TGGGINIPG------VLSSLPSLGFSLP-------TWGKVGLGLAGTMLLTPTCTLTI 264
                                                  :||
Db  321 RGEGVLMQGPGEKMLLVFLPTTSHRDPEDAQELLIHTAWNEL------KAVSPLCRPAA 373

Qy  265 RCCCCRRRCCGCN--------CCCRCC-----FCCR 287
            :||
Db  374 EALLCYHLFLECSPGVVPTPMPICREYCLAVKELFCAK 411
```

RESULT 5
I48697
protein-tyrosine kinase (EC 2.7.1.112) nsk2 precursor, splice form 1 - mouse
N;Alternate names: receptor-type tyrosine kinase
N;Contains: protein-tyrosine kinase nsk2 precursor, splice form 3
C;Species: Mus musculus (house mouse)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C;Accession: I48697; S60740
R;Ganju, P.; Walls, E.; Brennan, J.; Reith, A.D.
Oncogene 11, 281-290, 1995
A;Title: Cloning and developmental expression of Nsk2, a novel receptor tyrosine kinase
A;Reference number: I48697; MUID:95349951; PMID:7624144
A;Accession: I48697
A;Molecule type: DNA
A;Residues: 1-881 <GAN1>
A;Cross-references: EMBL:X86445; NID:g929725; PIDN:CAA60166.1; PID:g929726
A;Experimental source: splice form 1
A;Accession: S60740
A;Molecule type: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-456, 'A',466-881 <GAN2>
A;Cross-references: EMBL:X86445; NID:g929725
A;Experimental source: splice form 3

Right column

C;Comment: For alternate splice forms see PIR:I48696.

A;Gene: nsk2
A;Cross-references: MGI:103308
C;Genetics:
C;Superfamily: mouse ror-related receptor; immunoglobulin homology; protein kinase ho
C;Keywords: ATP; glycoprotein; phosphotransferase; receptor; transmembrane protein; t
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-881/Product: protein-tyrosine kinase nsk2, splice form 1 #status predicted <MAT1
F;22-456,'A',466-881/Product: protein-tyrosine kinase nsk2, splice form 3 #status pre
F;42-101/Domain: immunoglobulin homology <IMM1>
F;135-192/Domain: immunoglobulin homology <IMM2>
F;226-284/Domain: immunoglobulin homology <IMM3>
F;498-518/Domain: protein-tyrosine kinase homology predicted <TRM>
F;575-865/Domain: protein kinase homology <KIN>
F;583-591/Region: protein kinase ATP-binding motif
F;222,462/Binding site: carbohydrate (Asn) (covalent) #status predicted

```
Query Match        6.48;     Score 133.5;   DB 1;   Length 881;
Best Local Similarity   21.6%;   Pred. No. 0.046;
Matches  73;   Conservative  45;   Mismatches  127;   Indels  93;   Gaps  17;

Qy    4 GAMENRDPPGSGSGNEVIEGPQONARVLKGSQARFNCTVSOGWK-LIMWALSDMVVLSVRP 62
Db  113 GALQVKMKP------KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD--------- 158

Qy   63 MEPIITNDRFTSORYDQOGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-------- 117
Db  159 -NALRENSRIAALE-------SGSLRIHNVQKEDAGOYRCVAKNSL---GTAYSKLVKLEV 208

Qy  118 QVMGELFIPSVNLVVAAENEPCEVTC----LPSHWTWLPDISW-ELGLLVSHSSYFVPEP 172
Db  209 EVLGRILRAPESHNVTFGSFVTLRCTEIGIP----VPTISWIENGNAVSSGSIQESVKD 263

Qy  173 SDLQSAVSILALTPQSNGTLTCVAT-----WKSLKARKSATVNLTVIR---------CPQD 219
Db  264 RVIDSRLQLFITKP--GLYTCIATNKHGEKFSTAKAAATVSIAEWSKSQKDSQGYCAQY 320

Qy  220 TGGGINIPG------VLSSLPSLGFSLP-------TWGKVGLGLAGTMLLTPTCTLTI 264
Db  321 RGEGVLMQGPGEKMLLVFLPTTSHRDPEDAQELLIHTAWNEL------KAVSPLCRPAA 373

Qy  265 RCCCCRRRCCGCN--------CCCRCC-----FCCR 287
Db  374 EALLCYHLFLECSPGVVPTPMPICREYCLAVKELFCAK 411
```

RESULT 6
T24272
hypothetical protein T01B7.8 - Caenorhabditis elegans
R;Sims, M.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z19867
A;Accession: T24272
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-164 <WIL>
A;Cross-references: EMBL:Z66499; PIDN:CAA91301.1; GSPDB:GN000020; CESP:T01B7.8
A;Experimental source: clone T01B7
A;Gene: CESP:T01B7.8
C;Genetics:
A;Map position: 2
A;Introns: 20/3; 90/2

```
Query Match        6.28;     Score 129.5;   DB 2;   Length 164;
Best Local Similarity   32.1%;   Pred. No. 0.015;
Matches  43;   Conservative  11;   Mismatches  53;   Indels  27;   Gaps  6;

Qy  157 LGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRC 216
Db    6 LAILLAIGTFIAV----SQVQSAV----LPVSSTELATVGTDVDSTPASTAIDTLGNSSSRV 57
```

```
Qy  217 PQDTGGGINIPGVLSSLPSLGFSLPTWGRVGLGLAGTMLITPYCTLITRCCCCRRRCCGC 276
Db   58 KRQGGCGCGCGC---------------GCGCGGGGGG--CGCCCGRPRCCCCCRRCCTC 101

Qy  277 --NCCC-RCCFCCR 287
Db  102 CRTCCCTRCCTCCR 115
```

RESULT 7
T15651
hypothetical protein C27A2.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999
C;Accession: T15651
R;Nhan, M.
submitted to the EMBL Data Library, May 1996
A;Description:The sequence of C. elegans cosmid C27A2.
A;Reference number: Z18382
A;Accession: T15651
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-188 <NHA>
A;Cross-references: EMBL:U58760; NID:g1303384; PIDN:g1303389; PIDN:AAB00710.1; GSPDB:GN00
A;Experimental source: strain Bristol N2; clone C27A2
A;Genetics:
A;Gene: CESP:C27A2.5
A;Map position: 2
A;Introns: 19/3; 91/2

Query Match   5.9%;
Best Local Similarity   61.3%;
Matches   19; Conservative 1; Mismatches 8; Indels 3; Gaps 2;
Score 122.5; DB 2; Length 188;
Pred. No. 0.057;

```
Qy  260 CTLTIRCCCCRRRCCGC--NCCC-RCCFCCR 287
Db   86 CCCRPKCCCCRRCCTCCRTCCCTRCCTCCR 116
```

RESULT 8
I56551
neurotrimin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 19-May-2000
C;Accession: I56551
R;Struyk, A.F.; Canoll, P.D.; Wolfgang, M.J.; Rosen, C.L.; D'Eustachio, P.; Salzer, J.L.
J. Neurosci. 15, 2141-2156, 1995
A;Title: Cloning of neurotrimin defines a new subfamily of differentially expressed neur
A;Reference number: I56551; MUID:95198094; PMID:7891157
A;Accession: I56551
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-344 <RES>
A;Cross-references: EMBL:U16845; NID:g755184; PIDN:g755184; PIDN:AAA67445.1; PID:g755185
A;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin

Query Match   5.8%;
Best Local Similarity   24.1%;
Matches   59; Conservative 37; Mismatches 93; Indels 56; Gaps 13;
Score 121; DB 2; Length 344;
Pred. No. 0.14;

```
Qy   26 NARVLKGSQARFNCTVSQGWKLIMWALSDMVVLS------VRPMEPIITNDRFTSQRYDQ 79
Db   44 NVTVRQGESATLRCTIDNRVTRVAWLNRSTILYAGNDKWCLDPRVVLISN--TQTQY-- 98

Qy   80 GGNFTSEMIIHNVEPSDSGNIRCSLQ------NSRLHGSAYLTVQVMGELFIPSVNLVVA 133
Db   99 -----SIEIQNVDVDEGPYTCSVQTDNHPKTSRVH---LIVQVSPKIVEISDSISIN 148

Qy  134 ENEPCEVTCL----PSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSN 189
Db  149 EGNNISLTCIATGRPE-----PTVTWR--HISPKAVGFVSEDEYLE--IQGITREQS 196
```

```
Qy  190 GTLTCVATWKSLKARKSATVNLTVIRCP--------QDTGGGINIPGVL---SSLPSLGFSL 240
Db  197 GEYECSAS-NDVAAPVVRNVTNVTNYPPVIISEAKGTGPVVGQKGTLQCEASAVPSAEFQ- 254

Qy  241 PTWGK 245
Db  255 --WFK 257
```

RESULT 9
T29757
protein UNC-89 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C;Accession: T29757
R;Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A;Description:The sequence of C. elegans cosmid C09D1.
A;Reference number: Z20679
A;Accession: T29757
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6642 <DUZ>
A;Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
A;Experimental source: strain Bristol N2; clone C09D1
A;Genetics:
A;Gene: CESP:unc-89
A;Map position: 1
A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Query Match   5.7%;
Best Local Similarity   23.4%;
Matches   51; Conservative 39; Mismatches 93; Indels 35; Gaps 8;
Score 119.5; DB 2; Length 6642;
Pred. No. 4.3;

```
Qy   20 VIEGPQNARVLKGSQARFNCTVSQGW--KLIMWALSDMVVLSVRPMEPIITNDRFTSQRY 77
Db 2077 VVDGPKSVTIKETETAEFKAITS-GFPAPTVKWTINEKIVEESRIITTIKTEDVYT---- 2131

Qy   78 DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLIVQVMGELFIPS-----VNLVV 132
Db 2132 --------LKISNAKIEQTGTVKVTAQNSAGQDSKQADLKVEPNVKAPKFKSQLTDKVA 2182

Qy  133 AENEPCEVTCL---PSHWTWLPDISWEL-GLLVSHSSYYFVPEPSDLQSAVSILALTPQS 188
Db 2183 DEGEPLRWNLELDGPSPGT--EVSWLNGQPLTKSDTVQGVVDHGDGTYHVTIAEAKPEM 2239

Qy  189 NGTLTCVATWKSLKARKSATVNLT-------VIRCPQD 219
Db 2240 SGTLTAKAIKNAAGECETSAKYTVNGGNKRKPEFVQAPQN 2277
```

RESULT 10
JC4593
protein-tyrosine kinase-related receptor PTK7 precursor - human
N;Alternate names: receptor protein tyrosine kinase-like protein (RPTK)
C;Species: Homo sapiens (man)
C;Date: 16-Apr-1996 #sequence_revision 24-May-1996 #text_change 24-Sep-1999
C;Accession: JC4593
R;Park, S.K.; Lee, H.S.; Lee, S.T.
J. Biochem. 119, 235-239, 1996
A;Title: Characterization of the human full-length PTK7 cDNA encoding a receptor prot
A;Reference number: JC4593; MUID:97037064; PMID:8882711
A;Accession: JC4593
A;Molecule type: mRNA
A;Residues: 1-1070 <PAR>
A;Cross-references: GB:U40271; NID:g1322231; PIDN:AAC50484.1; PID:g1322232
A;Comment: This protein is a member of receptor protein tyrosine kinase family, but p
A;Genetics:
A;Gene: GDB:PTK7
A;Cross-references: GDB:134760; OMIM:601890
A;Map position: 6p21.1-6p12.2
A;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

```
Db  264 RVIDSRLQLFITKP---GLYTCIATNKHGEKFSTAKAAATISIAEWSKPQKDNNGYCAQY 320
Qy  224 ----INIPGVLSSLPSLGFSL------------PTWGKVGLGLAGTMLLTPTCTLITRCC 267
Db  321 RGEVCNAVLAKDALVFLNTSYADPEEAQELLVHTAWNEL------KVVSPVCRPAAEAL 373
Qy  268 CCRRRCCGCN---------CCRCC----FCCRRKRGFRIQFQKKSEKEKTNKETETE 312
Db  374 LCNHIFQECSPGVVPFIPICREYCLAVKELFCAKE-----WLVMEEKTHRGLYRSEMHLL 429
Qy  313 SGNENSGYNSDEQKTIDTASLPPKSCESSDPEQRNSSCGPPHQRADQRP----PRPASHP 368
Db  430 SVPECSKLPSMHWDPTACARLP----HLDYNKENLKTFPP--MTSSKPSVDIPNLPSSS 482
Qy  369 QASFNLA 375
Db  483 SSSFSVS 489
```

RESULT 4
US-08-644-271-29
; Sequence 29, Application US/08644271
; Patent No. 5814478
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
; TITLE OF INVENTION: AND LIGANDS
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/644,271
; FILING DATE: 10-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 60/008,657
; FILING DATE: 15-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert, Robert J
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 195A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; TELEX:
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 869 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-644-271-29

Query Match 6.3%; Score 132.5; DB 2; Length 869;
Best Local Similarity 20.8%; Pred. No. 0.0018;
Matches 89; Conservative 55; Mismatches 178; Indels 105; Gaps 20;

```
Qy    4 GAMENRDPGSGSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRP 62
Db  113 GALQVKMKP-----KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD------- 158
Qy   63 MEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMG 121
```

```
Db  159 -SPLRENSRIAVLE----------SGSLRIHNVQKEDAGQYRVAKNSL--GTAYSKVVKLEV 208
Qy  122 ELFIPSVNLVVAENEP--------CEVTCLPSHWTWLPDISW-ELGLLVSHSSYFVPEP 172
Db  209 EVFARILRAPESHNVTFGSFVTLHCTATGIP----VPTITWIENGNAVSSGSIOESVKD 263
Qy  173 SDLQSAVSILALTPQSNGTLTCVAT----WKSLKARKSATVNLTVIRCPQDTGGG---- 223
Db  264 RVIDSRLQLFITKP--GLYTCIATNKHGEKFSTAKAAATISIAEWSKPQKDNKGYCAQY 320
Qy  224 ----INIPGVLSSLPSLGFSL------------PTWGKVGLGLAGTMLLTPTCTLITRCC 267
Db  268 CCRRRCCGCN---------CCRCC----FCCRRKRGFRIQFQKKSEKEKTNKETETE 312
Db  321 RGEVCNAVLAKDALVFLNTSYADPEEAQELLVHTAWNEL------KVVSPVCRPAAEAL 373
Db  374 LCNHIFQECSPGVVPFIPICREYCLAVKELFCAKE-----WLVMEEKTHRGLYRSEMHLL 429
Db  430 SVPECSKLPSMHWDPTACARLP----HLDYNKENLKTFPP--MTSSKPSVDIPNLPSSS 482
Db  483 SSSFSVS 489
Qy  369 QASFNLA 375
```

RESULT 5
US-09-077-955-33
; Sequence 33, Application US/09077955A
; Patent No. 6413740
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/09/077,955A
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PCT/US96/20696
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 08/644,271
; EARLIER FILING DATE: 1996-05-10
; EARLIER APPLICATION NUMBER: 60/008,657
; EARLIER FILING DATE: 1995-12-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 33
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-077-955-33

Query Match 6.3%; Score 132.5; DB 4; Length 869;
Best Local Similarity 20.8%; Pred. No. 0.0018;
Matches 89; Conservative 55; Mismatches 178; Indels 105; Gaps 20;

```
Qy    4 GAMENRDPGSGSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRP 62
Db  113 GALQVKMKP-----KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD------- 158
Qy   63 MEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMG 121
Db  159 -SPLRENSRIAVLE----------SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKVVKLEV 208
Qy  122 ELFIPSVNLVVAENEP--------CEVTCLPSHWTWLPDISW-ELGLLVSHSSYFVPEP 172
Db  209 EVFARILRAPESHNVTFGSFVTLHCTATGIP----VPTITWIENGNAVSSGSIQESVKD 263
Qy  173 SDLQSAVSILALTPQSNGTLTCVAT----WKSLKARKSATVNLTVIRCPQDTGGG---- 223
Db  264 RVIDSRLQLFITKP---GLYTCIATNKHGEKFSTAKAAATISIAEWSKPQKDNKGYCAQY 320
Qy  224 ----INIPGVLSSLPSLGFSL------------PTWGKVGLGLAGTMLLTPTCTLITRCC 267
```

```
Db  321 RGEVCNAVLAKDALVPLNTSYADPEEAQELLVHTAWNEL------------KVVSPVCRPAAEAL 373
         |:|               :.|:            |                ::|:|  |
Qy  268 CCRRRCCGCN------CCCRCC----FCCRRKRGFRIQFQKKSEKEKINKETETE 312

Db  374 LCNHIFQECSPGVYPTPIPICREYCLAVKELFCAKE----WLVMEEKTHRGLYRSEMILL 429
Qy

Db  313 SGNENSGYNSDEQKTTDTASLPPKSCESSDPEQRNSSCGPPHQRADQRP----PRPASHP 368
Qy

Db  430 SVPECSKLPSMHWDPTACARLP-----HLDYNKENLKTFPP--MTSSKPSVDIPNLPSSS 482
Qy

Db  369 QASFNLA 375
Qy  483 SSSFSVS 489
```

RESULT 6
US-08-977-767-3
Sequence 3, Application US/08977767
Patent No. 5972684
GENERAL INFORMATION:
 APPLICANT: Bandman, Olga
 APPLICANT: Yue, Henry
 APPLICANT: Greenwald, Sara
 APPLICANT: Corley, Neil C.
 TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
  ADDRESSEE: Incyte Pharmaceuticals, Inc.
  STREET: 3174 Porter Drive
  CITY: Palo Alto
  STATE: CA
  COUNTRY: USA
  ZIP: 94304
 COMPUTER READABLE FORM:
  MEDIUM TYPE: Diskette
  COMPUTER: IBM Compatible
  OPERATING SYSTEM: DOS
  SOFTWARE: FastSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/977,767
  FILING DATE: Herewith
  CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
  APPLICATION NUMBER:
  FILING DATE:
 ATTORNEY/AGENT INFORMATION:
  NAME: Billings, Lucy J.
  REGISTRATION NUMBER: 36,749
  REFERENCE/DOCKET NUMBER: PF-0423 US
 TELECOMMUNICATION INFORMATION:
  TELEPHONE: 650-855-0555
  TELEFAX: 650-845-4166
  TELEX:
INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
  LENGTH: 1345 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
 IMMEDIATE SOURCE:
  LIBRARY: GenBank
  CLONE: 1532042

```
Query Match    6.2%;   Score 128.5;   DB 2;   Length 1345;
Best Local Similarity  33.0%;   Pred. No. 0.0077;
Matches  37;  Conservative  1;  Mismatches  41;  Indels  33;  Gaps  5;

Qy  190 GTLTCVATWKSLRARKSATVNLJTVIRCPQDTGGGI----NIPGVLSSLPSLGFSLPTWGK 245
        |||||                        |||  | |||
Db  414 GTCTCTGT-------------------GC-CGTGGGAAGCGTCAGAGCCCCGTGGATGTGGA 455
```

```
Qy  246 VGLGLAGTMLLT-PTCTLTIRCCCCRRRCCGCNCCCRC------------CFCC 286
Db  456 CGTGAAGAGGGTGCTCCTCTATGACCCCTTCGTTGTCCCCCTCTGCGACACTCAGCACC 507
```

RESULT 7
US-09-651-200-2
Sequence 2, Application US/09651200
Patent No. 6429303
GENERAL INFORMATION:
 APPLICANT: Green et al.
 TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
 TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
 TITLE OF INVENTION: Polypeptides Encoded Thereby
 FILE REFERENCE: 15966-562 (CURA-62)
 CURRENT APPLICATION NUMBER: US/09/651,200
 CURRENT FILING DATE: 2000-08-30
 PRIOR APPLICATION NUMBER: 60/152383
 PRIOR FILING DATE: 2000-08-30
 PRIOR APPLICATION NUMBER: 60/152383
 PRIOR FILING DATE: 1999-09-03
 PRIOR APPLICATION NUMBER: 60/172909
 PRIOR FILING DATE: 1999-12-21
 PRIOR APPLICATION NUMBER: 60/183578
 PRIOR FILING DATE: 2000-02-18
 NUMBER OF SEQ ID NOS: 25
 SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
 LENGTH: 340
 TYPE: PRT
 ORGANISM: Homo sapiens
US-09-651-200-2

```
Query Match    5.98;   Score 123.5;   DB 4;   Length 340;
Best Local Similarity  21.3%;   Pred. No. 0.0033;
Matches  78;  Conservative  50;  Mismatches  130;  Indels  109;  Gaps  17;

Qy  11 PPGSGSGNEVIEGPQNARV-LKGSQARFNCIVS--QGWKL---IMWALSDMVLSVRPM 63
       |  ||| ||                |||                  ::|:|  |
Db  45 PQRSPTGAVEVQVPEDPVVALNGTDATLHCSFSPEPGFSLTQLNLIWQLTDTKQLV---- 100

Qy  64 EPIITNDRFTSQRYDGGGN. ------------TSEMIHNVEPSDSGNIRCSLQNSRL 109
Db  101 ------HSFTEGR-DQGSAYANRTALFPDLIAQGNASLRLQRVRVADEGSFTCFV-SIRD 152

Qy  110 HGSAYLIVQVMGELFIPSVNLV-----VAENEPCEVTICLPSHWTWLP--DISWELGL---L 160
Db  153 FGSAAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPL 210

Qy  161 VSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP--- 217
Db  211 TGNVTLSQMANEQGLFDVHSVLRVVLGANGTYSC---------LVRNPVLQ 252

Qy  218 QDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCCRRRCCGCN 277
Db  253 QDAHGSVTIIGQPMIFPPEAL----WVIVGLSVCLIALLV------------ 288

Qy  278 CCCRCCFCCRRKRGFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKTTDTASLPPKS 337
Db  289 ---ALAFVCWRK-------IKQSCEENAGAEDQDG------EGEGSKTALQPLKH 328

Qy  338 CESSDPE 344
Db  329 SDSKEDD 335
```

RESULT 8
US-09-651-200-4
Sequence 4, Application US/09651200
Patent No. 6429303
GENERAL INFORMATION:
 APPLICANT: Green et al
 TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
 TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
 TITLE OF INVENTION: Polypeptides Encoded Thereby

```
FILE REFERENCE: 15966-562 (CURA-62)
CURRENT APPLICATION NUMBER: US/09/651,200
CURRENT FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/152383
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/172909
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/183578
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 4
LENGTH: 441
TYPE: PRT
ORGANISM: Homo sapiens
US-09-651-200-4

Query Match    5.9%;  Score 123.5;  DB 4;  Length 441;
Best Local Similarity   21.3%;  Pred. No. 0.0047;
Matches  78;  Conservative  50;  Mismatches 130;  Indels 109;  Gaps 17;

Qy   11 PPGSGSGNEVIEGPQNARV-LKGSQARFNCTVS--QGWKL---IMWALSDMVVLSVRPM 63
          ||   : |||   : | :||  :|   |: |||   |||
Db  146 PQRSPTGAVEVQVPEDPVVALVGTDATLHCSFSPEPGFSLTQLNLIWQLTDTKQLV---- 201

Qy   64 EPIITNDRFTSQRYDQGGNF------------TSEMIIHNVEPSDSGNIRCSLQNSRL 109
               :||         |||  :|| :  | ||
Db  202 ----HSFTEGR-DQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFV-SIRD 253

Qy  110 HGSAYLTVQVMGELFIPSVNLV------VAENEPCEVTCLPSHWTWLP--DISWELGL--L 160
         |  :  |  : | ||  |  : |       |  |: || |
Db  254 FGSAAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPL 311

Qy  161 VSHSSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP--- 217
          : |:  | |: |||   : || : |:|        | :|        :|| :|
Db  312 TGNVTTSQMANEQGLFDVHSVLRVVLGANGTYSC---------------LVRNPVLQ 353

Qy  218 QDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCRRRCCGCN 277
          || ::  |||  :| |||  :||  |||||   |||
Db  354 QDAHGSVIITGQPMTFPPEAL-----WVTVGLSVCLIALLV------------------- 389

Qy  278 CCCRCCFCCRRKRGFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKTTDFASLPPKS 337
          | ||    ||  |                          | ||
Db  390 ---ALAFVCWRK------------IKQSCEEENAGAEDQDG-----EGEGSKTALQPLKH 429

Qy  338 CESSDPE 344
         : |:
Db  430 SDSKEDD 436
```

```
ORGANISM: Homo sapiens
US-09-651-200-6

Query Match    5.9%;  Score 122.5;  DB 4;  Length 534;
Best Local Similarity   21.3%;  Pred. No. 0.0075;
Matches  78;  Conservative  49;  Mismatches 131;  Indels 109;  Gaps 17;

Qy   11 PPGSGSGNEVIEGPQNARV-LKGSQARFNCTVS--QGWKL---IMWALSDMVVLSVRPM 63
          ||   : |||   : | :||  :|   |: |||   |||
Db  239 PQRSPTGAVEVQVPEDPVVALVGTDATLRCSFSPEPGFSLAQLNLIWQLTDTKQLV---- 294

Qy   64 EPIITNDRFTSQRYDQGGNF------------TSEMIIHNVEPSDSGNIRCSLQNSRL 109
               :||         |||  :|| :  | ||
Db  295 ----HSFTEGR-DQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFV-SIRD 346

Qy  110 HGSAYLTVQVMGELFIPSVNLV------VAENEPCEVTCLPSHWTWLP--DISWELGL--L 160
         |  :  |  : | ||  |  : |       |  |: || |
Db  347 FGSAAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPL 404

Qy  161 VSHSSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP--- 217
          : |:  | |: |||   : || : |:|        | :|        :|| :|
Db  405 TGNVTTSQMANEQGLFDVHSVLRVVLGANGTYSC---------------LVRNPVLQ 446

Qy  218 QDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCRRRCCGCN 277
          || ::  |||  :| |||  :||  |||||   |||
Db  447 QDAHGSVIITGQPMTFPPEAL-----WVTVGLSVCLIALLV------------------- 482

Qy  278 CCCRCCFCCRRKRGFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKTTDFASLPPKS 337
          | ||    ||  |                          | ||
Db  483 ---ALAFVCWRK------------IKQSCEEENAGAEDQDG-----EGEGSKTALQPLKH 522

Qy  338 CESSDPE 344
         : |:
Db  523 SDSKEDD 529
```

```
RESULT 9
US-09-651-200-6
Sequence 6, Application US/09651200
Patent No. 6429303
GENERAL INFORMATION:
APPLICANT: Green et al
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
TITLE OF INVENTION: Polypeptides Encoded Thereby
FILE REFERENCE: 15966-562 (CURA-62)
CURRENT APPLICATION NUMBER: US/09/651,200
CURRENT FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/152383
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/172909
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/183578
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 6
LENGTH: 534
TYPE: PRT
```

```
RESULT 10
US-09-651-200-24
Sequence 24, Application US/09651200
Patent No. 6429303
GENERAL INFORMATION:
APPLICANT: Green et al
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
TITLE OF INVENTION: Polypeptides Encoded Thereby
FILE REFERENCE: 15966-562 (CURA-62)
CURRENT APPLICATION NUMBER: US/09/651,200
CURRENT FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/152383
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/172909
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/183578
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 24
LENGTH: 534
TYPE: PRT
FEATURE:
ORGANISM: Unknown
OTHER INFORMATION: Description of Unknown Organism: Sequence
OTHER INFORMATION: mz5020.protein from Figure 4.
US-09-651-200-24

Query Match    5.9%;  Score 122.5;  DB 4;  Length 534;
Best Local Similarity   21.3%;  Pred. No. 0.0075;
Matches  78;  Conservative  49;  Mismatches 131;  Indels 109;  Gaps 17;

Qy   11 PPGSGSGNEVIEGPQNARV-LKGSQARFNCTVS--QGWKL---IMWALSDMVVLSVRPM 63
          ||   : |||   : | :||  :|   |: |||   |||
Db  239 PQRSPTGAVEVQVPEDPVVALVGTDATLRCSFSPEPGFSLAQLNLIWQLTDTKQLV---- 294
```

RESULT 11

US-09-062-365-1

; Sequence 1, Application US/09062365
; Patent No. 6465422
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Ann Marie
; APPLICANT: Stern, David
; TITLE OF INVENTION: METHOD FOR INHIBITING TUMOR INVASION OR SPREADING IN A
; TITLE OF INVENTION: SUBJECT
; FILE REFERENCE: 55424
; CURRENT APPLICATION NUMBER: US/09/062,365
; CURRENT FILING DATE: 1998-04-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 1
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Human
US-09-062-365-1

Query Match   5.8%;   Score 121.5;   DB 4;   Length 332;
Best Local Similarity 25.7%;   Pred. No. 0.0048;
Matches 57;   Conservative 27;   Mismatches 69;   Indels 69;   Gaps 11;

```
Qy   64 EPIITNDRFTSQRYDQGGNF-------TSEMIHNVEPSDSGNIRCSLQNSRL 109
Db  295 ------HSFTEGR-DQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFV-SIRD 346

Qy  110 HGSAYLTVQVMGELFIPSVNLV------VAENEPCEVTCLPSHWTWLP--DISWELGL--L 160
Db  347 FGSAAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPL--L 404

Qy  161 VSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP--- 217
Db  405 TGNVTTSQMANEQLFDVHSVLRVVLGANGTYSC-------LVRNPVLQ 446

Qy  218 QDTGGGINIPGVLSLSPLGFSLPTWGKVGLGLAGTMLLTPTCTLFIRCCCRRCCGCN 277
Db  447 QDAIGSVITGQPMTFPPEAL------WVTVGLSVCLIALLV 482

Qy  278 CCCRCCFCCRRKRGFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKTDTASLPPKS 337
Db  483 ---ALAFVCWRK---------------IKQSCEENAGAEDQDG 522

Qy  338 CESSDPE 344
Db  523 SDSKEDD 529
```

RESULT 12

US-09-540-245A-15

; Sequence 15, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kidd, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions
; FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 15
; LENGTH: 1395
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-540-245A-15

Query Match   5.7%;   Score 120;   DB 4;   Length 1395;
Best Local Similarity 26.2%;   Pred. No. 0.047;
Matches 66;   Conservative 29;   Mismatches 101;   Indels 56;   Gaps 15;

```
Qy   24 PQNARVLKGSQARFNCTVSQG--WKLIMV------ALSDMVVLSVRPMEPIITNDRFTSQR 76
Db  161 PKDTRVAKGETALLECGPPKGIPEPTLIWIKDGVPLDDLKAMSFGASSRV--------R 211

Qy   77 YDQGGNFTSEMIHNVEPSDSGNIRCSLQN---SRLHGSAYLTVQVMGELFIPSVNLVVA 133
Db  212 IVDGGN---LIISNVEPIDEGNYKCIAQNLVGTRESSYAKLIVQVKPYPMKEPKDQVML 267

Qy  134 ENEP----CEVTCLPSHWTWLPDISW--ELGLLVSHSSYYFVPEPSDLQSAVSILALTP 186
Db  268 YGQTATFHCSVGGDP----PKVLMKKEEGNIPVSRARILHD------EKSLEISNITP 316

Qy  187 QSNGTLTCVA--TWKSLKARKSATV----NLTVIRCPQDTGGGINIPGVLSLSPLGFSL 240
Db  317 TDEGTYVCEAHNNVGQISARASLIVHAPPNFT--KRPSNKKVGLN--GVV-QLPCMASGN 371

Qy  241 PT----WGKVGL 248
Db  372 PPPSVFWTKEGV 383
```

Matches 57;   Conservative 27;   Mismatches 69;   Indels 69;   Gaps 11;

```
Qy   64 EPIITNDRFTS-----QRYDQGGNFT--SEMIHNVEPSDSGNIR----CSLLQNSRLHGS 112
Db  140 KPLVPNEKGVSVKEQTRRHPETGLPTLLQSSELM---VTPARGGDPRPTPSCSFSPGLPHR 196

Qy  113 AYLTVQVMGELFIP----------SVNLVVAENEP------CEVTCLPSHWTWLPDIS 154
Db  197 ALRTAPIQPRWEPVPLEEVQLVV---EPEGGAVAPGGTVTLTCEVPAQPS----PQIH 248

Qy  155 WELGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVI 214
Db  249 WMKD-------GVPLPLPPSPVLILPEIGPDQGTYSCVATHSSHGPQESRAVSISII 299

Qy  215 RCPQDTGGGINIPGVLSLSPLGFSLPTWGKVGLGLAGTMLL 256
Db  300 E-PGEEG---------PTAGSVGGSGLGTLAL 321
```

RESULT 13

US-08-374-834-1

; Sequence 1, Application US/08374834
; Patent No. 5655473
; GENERAL INFORMATION:
; APPLICANT: Valenzuela, et al.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTOR
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill River Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,834
; FILING DATE: 19-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/095,658
; FILING DATE: 21-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert, Robert J.
; REGISTRATION NUMBER: 36,108

REFERENCE/DOCKET NUMBER: REG 190A
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (914) 345-7400
  TELEFAX: (914) 345-7721
REFERENCE FOR SEQ ID NO: 1:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
  LENGTH: 868 amino acids
  TYPE: amino acid
  STRANDEDNESS:
  TOPOLOGY: unknown
  MOLECULE TYPE: protein
US-08-374-834-1

Query Match   5.7%; Score 118; DB 1; Length 868;
Best Local Similarity 21.0%; Pred. No. 0.037;
Matches 83; Conservative 42; Mismatches 151; Indels 120; Gaps 19;

```
Qy    4 GAMENRDPPGSGSGNEVIEGPQNARVLKGSQARFNGTVSQGWK-LIMWALSDMVVLSVRP 62
          ::  :  :     :          ::: : :: :  :::: ::: :::   :
Db  113 GALQVKMKP-----KITRPIINVKIIEGLKAVLPCTTMGNPKPSVSWIKGDSALRE--- 163

Qy   63 MEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMG 121
          : :  :   :    :: ::   :::  : :  : :: : : :: : ::: : : : :
Db  164 -----NSRIAVLE---SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKLVKLEV 208

Qy  122 ELFIPSVNLVVAENEP------CEVTCLPSHWTWLPDISW-ELGLLVSHSSYYFVPEP 172
         : :  :  : :       : :   :  :: ::  ::  : : ::
Db  209 EVFARILRAPESHNVTEGFSVYLRCTAIGMP----VPTISWIENGNAVSSGSIQENVKD 263

Qy  173 SDLQSAVSILALTPQSNGTLTCVAT-----WKSLKARKSATVNLTVIR--------CPQD 219
         : :  :  :   : :  ::       :: :   :::  ::
Db  264 RVIDSRLQLFITKP--GLYTCIATNKHGEKFSTAKAAATVSIAEWSKSQKESKGYCAQY 320

Qy  220 TGGGINIPGVLSSLPSLGFSLP--------TWGKVGLGLAGTMLLTPTCTLTIRCC 267
         : :  :  :          :: : :  :::  ::
Db  321 RGEVCDAVLVKDSLVFFNTSYPDPEEAQELLIHTAWNEL-----KAVSPLCRPAAEAL 373

Qy  268 CCRRRCCGCN-------------CCCRCC--------FCCRRKRGFRIQFQKKSEKETNKETETE 312
         :  :                :::           ::
Db  374 LCNHLFQECSPGVLPTMPICREYCLAVKELFCA------------KEWLAMEGKTH 418

Qy  313 SGNENSGYYNSDEQKTTDJTASLPPKSCES----SDP 343
         :
Db  419 RGLYRSGMH---------FLPVPECSKLPSMHQDP 444
```

RESULT 14
US-08-644-271-1
; Sequence 1, Application US/08644271
; Patent No. 5814478
GENERAL INFORMATION:
APPLICANT: Valenzuela, et al.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/644,271
FILING DATE: 10-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/008,657

FILING DATE: 15-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
  LENGTH: 868 amino acids
  TYPE: amino acid
  STRANDEDNESS:
  TOPOLOGY: unknown
  MOLECULE TYPE: protein
US-08-644-271-1

Query Match   5.7%; Score 118; DB 2; Length 868;
Best Local Similarity 21.0%; Pred. No. 0.037;
Matches 83; Conservative 42; Mismatches 151; Indels 120; Gaps 19;

```
Qy    4 GAMENRDPPGSGSGNEVIEGPQNARVLKGSQARFNGTVSQGWK-LIMWALSDMVVLSVRP 62
          ::  :  :     :          ::: : :: :  :::: ::: :::   :
Db  113 GALQVKMKP-----KITRPIINVKIIEGLKAVLPCTTMGNPKPSVSWIKGDSALRE--- 163

Qy   63 MEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMG 121
          : :  :   :    :: ::   :::  : :  : :: : : :: : ::: : : : :
Db  164 -----NSRIAVLE---SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKLVKLEV 208

Qy  122 ELFIPSVNLVVAENEP------CEVTCLPSHWTWLPDISW-ELGLLVSHSSYYFVPEP 172
         : :  :  : :       : :   :  :: ::  ::  : : ::
Db  209 EVFARILRAPESHNVTEGFSVYLRCTAIGMP----VPTISWIENGNAVSSGSIQENVKD 263

Qy  173 SDLQSAVSILALTPQSNGTLTCVAT-----WKSLKARKSATVNLTVIR--------CPQD 219
         : :  :  :   : :  ::       :: :   :::  ::
Db  264 RVIDSRLQLFITKP--GLYTCIATNKHGEKFSTAKAAATVSIAEWSKSQKESKGYCAQY 320

Qy  220 TGGGINIPGVLSSLPSLGFSLP--------TWGKVGLGLAGTMLLTPTCTLTIRCC 267
         : :  :  :          :: : :  :::  ::
Db  321 RGEVCDAVLVKDSLVFFNTSYPDPEEAQELLIHTAWNEL-----KAVSPLCRPAAEAL 373

Qy  268 CCRRRCCGCN-------------CCCRCC--------FCCRRKRGFRIQFQKKSEKETNKETETE 312
         :  :                :::           ::
Db  374 LCNHLFQECSPGVLPTMPICREYCLAVKELFCA------------KEWLAMEGKTH 418

Qy  313 SGNENSGYYNSDEQKTTDJTASLPPKSCES----SDP 343
         :
Db  419 RGLYRSGMH---------FLPVPECSKLPSMHQDP 444
```

RESULT 15
US-09-077-955-1
; Sequence 1, Application US/09077955A
; Patent No. 6413740
GENERAL INFORMATION:
APPLICANT: Valenzuela et al., David M.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
FILE REFERENCE: REG195-B-PCT-US
CURRENT APPLICATION NUMBER: US/09/077,955A
CURRENT FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: PCT/US96/20696
EARLIER FILING DATE: 1996-12-13
EARLIER APPLICATION NUMBER: 08/644,271
EARLIER FILING DATE: 1996-05-10
EARLIER APPLICATION NUMBER: 60/008,657
EARLIER FILING DATE: 1995-12-15
NUMBER OF SEQ ID NOS: 36
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 868
TYPE: PRT
ORGANISM: Rattus sp.

US-09-077-955-1

```
Query Match          5.7%;    Score 118;   DB 4;   Length 868;
Best Local Similarity   21.0%;    Pred. No. 0.037;
Matches   83;  Conservative   42;  Mismatches 151;  Indels 120;  Gaps  19;

Qy     4 GAMENRDPPGSGSGGNEVIEGPQNARVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRP  62
         |||:||| ||:   ||   |:|:||::||:|| || :||:||||| |:|||  |:| ||||
Db   113 GALQVKMKP-----------KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGDSALRE---  163

Qy    63 MEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMG  121
         :  ||||| | |||  | |||:||  ||| ||:|| |:|||||| |:| ||::
Db   164 ------NSRIAVLE-----SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKLVKLEV  208

Qy   122 ELFIPSVNLVVAENEP----------CEVTCLPSHWTWLPDISW-ELGLLVSHSSYFVPEP  172
         |:|:|   | |          |:|  |   ||  ||:||| | ||: |
Db   209 EVFARILRAPESHNVTFGSSFVTLRCTAIGMP----VPTISWIENGNAVSSGSIQENVKD  263

Qy   173 SDLQSAVSTLALTPQSNGTLTCVAT-----WKSLKARKSATVNLTVIR--------CPQD  219
         :  ::| | | |: |      |   |||:  | ||: |                |:||
Db   264 RVIDSRLQLFITKP----GLYTCIATNKHGEKFSTAKAAATVSIAEWSKSQKESKGYCAQY  320

Qy   220 TGGGINIPGVLSSLPSILGFSLP-------TWGKVGLGLAGTMLLTPTCTLTIRCC  267
         |:|    | |      |:|    |||::   : ||  |
Db   321 RGEVCDAVLVKDSLVFENTSYPDPEEAQELLIHTAWNEL----KAVSPLCRPAAEAL  373

Qy   268 CCRRRCCGCN------CCCRCC------FCCCRKRGFRIQFQKKSEKEKTNKETETE  312
         ||:||               ||:|                   ||
Db   374 LCNHLFQECSPGVLPTMPICREYCLAVKELFCA-----------------KEWLAMEGKTH  418

Qy   313 SGNENSGYNSDEQKTLDTASLPPKSCES-----SDP  343
         |
Db   419 RGLYRSGMH----------FLFVPECSKLFSMHQDP  444
```

Search completed: April 28, 2003, 21:12:19
Job time : 19.0485 secs

OM protein - protein search, using sw model

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 .Compugen Ltd.

Run on: April 28, 2003, 16:24:31 : Search time 38.4662 Seconds
                       (without alignments)
                       1337.141 Million cell updates/sec

Title: US-09-729-264-4
Perfect score: 2088
Sequence: 1 MVAGAMENRDPPGSGSGNEV.........HPQASFNLASPEKVSNTTVV 386

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Post-processing:
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :   A_Geneseq_101002:*
               Minimum Match 0%
               Maximum Match 100%
               Listing first 45 summaries

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2088	100.0	386	23	AAU75541	Human B7-like prot
2	2004	96.0	382	23	AAU75540	Human B7-like prot
3	2000	95.8	386	23	AAU75542	Human B7-like prot
4	1576	75.5	377	23	AAU75543	Human B7-like prot
5	1318	63.1	463	22	ABG28169	Novel human diagno
6	913.5	43.8	370	23	AAU75544	Mouse B7-like prot
7	572.5	27.4	631	23	AAU75547	Rat B7-like protei
8	566.5	27.1	270	23	AAU75545	Mouse B7-like prot
9	485	23.2	223	23	AAU75546	Mouse B7-like prot
10	145.5	7.0	404	22	AAB81925	Extracorporeal clr

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

11	145.5	7.0	404	23	AAE23219	Human receptor for
12	145.5	7.0	404	23	AAU75543	Human receptor for
13	145.5	7.0	404	23	AAM48745	Human RAGE protein
14	138	6.6	1496	21	AAW81030	Melanoma associate
15	138	6.6	1498	22	AAY70469	Human p53 target m
16	138	6.6	1447	16	ABB11587	Human peroxidasin
17	136	6.5	1447	20	AAR68553	Deleted in colorec
18	136	6.5	1447	20	AAY33498	Human DCC protein.
19	136	6.5	1728	12	AAB50693	Human UNC-40 prote
20	136	6.5	1447	22	AAR13144	Deleted in Colorec
21	133.5	6.4	467	19	AAW62575	Nsk2 extracellular
22	133.5	6.4	475	17	AAR94982	Alternatively spli
23	133.5	6.4	863	19	AAW62569	Mouse receptor tyr
24	133.5	6.4	867	19	AAR84087	Nsk2 receptor. Mu
25	133.5	6.4	871	17	AAR62583	Mouse receptor tyr
26	133.5	6.4	871	17	AAW62568	Mouse receptor tyr
27	133.5	6.4	873	17	AAR84092	Nsk2 receptor with
28	133.5	6.4	873	19	AAW62573	Alternatively spli
29	133.5	6.4	881	17	AAR84091	Nsk2 receptor with
30	133.5	6.3	869	19	AAW62572	Mouse Nsk2 (altern
31	132.5	6.3	881	19	AAW26611	Mouse muscle-speci
32	132.5	6.3	869	18	AAW26506	Human Dmk receptor
33	131.5	6.3	537	22	AAG77857	Human receptor mMu
34	131.5	6.3	576	22	AAG77858	Mutant protein mMu
35	131.5	6.3	869	22	AAG77856	Protein of muscle
36	131.5	6.3	869	22	AAB68421	Protein of muscle
37	130.5	6.2	1483	22	ABG16336	Novel human diagno
38	128	6.1	592	23	ABB75751	Human immunoglobul
39	128	6.1	594	23	ABB75753	Human pancreas GP3
40	128	6.1	576	22	ABB75752	Novel human diagno
41	128	6.1	708	23	AAU29315	Human gp354 (putat
42	126	6.0	1953	23	AAU84351	Protein MYlk diffe
43	125.5	6.0	467	17	AAR84004	Nsk2 receptor with
44	123.5	5.9	340	23	AAU00904	Human B lymphocyte
45	123.5	5.9	441	22	AAU00905	Human B lymphocyte

RESULT 1
AAU75541

AAU75541 standard; Protein; 386 AA.

ALIGNMENTS

ID   AAU75541 standard; Protein; 386 AA.
XX
AC   AAU75541;
XX
DT   23-APR-2002 (first entry)
XX
DE   Human B7-like protein, B7-L_h2.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Homo sapiens.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;

```
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13029.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 2; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I).
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of human B7-L_h2.
XX
SQ   Sequence   386 AA;

Query Match             100.0%;  Score 2088;  DB 23;  Length 386;
Best Local Similarity   100.0%;
Matches  386;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1  MVAGAMENRDPPGSGSGNEVIEGPQNARVLKGSQARPNCTYSQGWKLIMWALSDMVVLSV   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MVAGAMENRDPPGSGSGNEVIEGPQNARVLKGSQARPNCTYSQGWKLIMWALSDMVVLSV   60

Qy     61  RPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAVLTVQVM  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  RPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAVLTVQVM  120

Qy    121  GELFIPSVNLVVAENEPCEVVCLPSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVS  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  GELFIPSVNLVVAENEPCEVVCLPSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVS  180

Qy    181  ILALTPQSNGITCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLGFSL     240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  ILALTPQSNGITCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLGFSL     240

Qy    241  PTWGKVGLGLAGTMLLITPCTLTIRCCCRRRCGCGNCCRCRCPCCRRKRGFRIQFQKKS  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  PTWGKVGLGLAGTMLLITPCTLTIRCCCRRRCGCGNCCRCRCPCCRRKRGFRIQFQKKS  300

Qy    301  EKEKTNKETETESGNENSGYNSDEQKTTDTDTASLPFKSCESSDPEQRNSSCGPPHQRQDR  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  EKEKTNKETETESGNENSGYNSDEQKTTDTDTASLPPKSCESSDPEQRNSSCGPPHQRADQR  360

Qy    361  PPRPASHPQQASFNLASPEKVSNTTVV  386
           ||||||||||||||||||||||||||||
Db    361  PPRPASHPQQASFNLASPEKVSNTTVV  386

RESULT 2
ID   AAU75540 standard; Protein; 382 AA.
XX
AC   AAU75540;
XX
DT   23-APR-2002  (first entry)
XX
DE   Human B7-like protein, B7-L_h1.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Homo sapiens.
XX
PN   WO2002007l0-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13028.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 1; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I).
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
```

```
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of human B7-Lh1.
XX
SQ   Sequence   382 AA;

Query Match           96.0%;  Score 2004;  DB 23;  Length 382;
Best Local Similarity   99.5%;  Pred. No. 8e-158;
Matches  372;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    13 GSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72
         |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db     9 GSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 68

Qy    73 TSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV 132
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69 TSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV 128

Qy   133 AENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTL 192
         |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   129 AENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTIL 188

Qy   193 TCVATWKSLKARKSAATVNLTVTRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGIAG 252
         ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   189 TCVATWKSLKARKSATVNLTVTRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGIAG 248

Qy   253 TMLLTPTCTLTIRCCCRRRCCGCNCCCRCCPCCRRKRGRFRIQFQKKSEKEKTNKETETE 312
         ||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db   249 TMLLTPTCTLTIRCCCRRRCCGCNCCCFCCRRKRGFRIQFQKKSEKEKTNKETETE 308

Qy   313 SGNENSGYNSDEQKTTDPASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASF 372
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   309 SGNENSGYNSDEQKTTDPASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASF 368

Qy   373 NLASPEKVSNTTVV 386
         ||||||||||||||
Db   369 NLASPEKVSNTTVV 382
```

```
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-L polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of human B7-L_h3.
XX
SQ   Sequence   386 AA;

Query Match           95.8%;  Score 2000;  DB 23;  Length 386;
Best Local Similarity   99.2%;  Pred. No. 1.7e-157;
Matches  371;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    13 GSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72
         |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    13 GSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72

Qy    73 TSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV 132
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    73 TSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV 132

Qy   133 AENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTL 192
```

RESULT 4

```
ID   AAU75543 standard; protein; 377 AA.
AC   AAU75543;
XX
DT   23-APR-2002  (first entry)
XX
DE   Human B7-like protein, B7-L_h4.
XX
KW   B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Homo sapiens.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13031.
XX
PT   New B7-like polypeptides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 4; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
```

```
Qy    373  NLASPEKVSNTIVV 386
           ||||||||||||||
Db    373  NLASPEKVSNTIVV 386

Db    313  SGNENSGYNSDEQKTTETPASLLPPKSCESSDPEQRNSSCGPPHQRADQRPPASHPQASF 372
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    313  SGNENSGYNSDEQKTTDTASLLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASF 372

Db    253  TMLLTPTCLTLIRCCCCRRRCCGCNCCCRRCCPCCRRKRGFRIQFQKKSEKEKTNKETETE 312
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    253  TMLLTPTCLTLIRCCCCRRRCCGCNCCCRRCCPCCRRKRGFRIQFQKKSEKEKTNKETETE 312

Qy    253  TMLLTPTCLTLIRCCCCRRRCCGCNCCCRRCCPCCRRKRGFRIQFQKKSEKEKTNKETETE 312

Db    193  TCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG 252
Qy    193  TCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG 252

Db    133  AENEPCEVTIGLPSHWTRLPDISWELGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTL 192
```

RESULT 5

```
ID   ABG28169 standard; protein; 463 AA.
AC   ABG28169;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #28160.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
XX
SQ   Sequence  377 AA;
```

```
Query Match             75.5%;  Score 1576;  DB 23;  Length 377;
Best Local Similarity   88.6%;
Matches  296;  Conservative  12;  Mismatches  20;  Indels   6;  Gaps  2;

Qy    1  MVAGAMENRDPPGSSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSV 60
         |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MVAGAMENRDPPGSGSGGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSV 60

Qy   61  RPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVM 120
         |||||||||||||||||||||| ||||||||||||||||||||||||||||||||| |||
Db   61  RPMEPIITNDRFTSQRYDQGGNLTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLIVQVM 120

Qy  121  GELFIPSVNLVVAENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYFVPEPSDLQSAVS 180
         ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  121  GELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYFVPEPSDLQSAVS 180

Qy  181  ILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSL 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  ILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSL 240

Qy  241  PTWGKVGLGLAGTMLLTPTCLTLIRCCCCRRRCCGCNCCCRRCCPCCRRKRGFRIQFQKKS 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||     ::
Db  241  PTWGKVGLGLAGTMLLTPTCLTLIRCCCCRRRCCGCNCCCRRCCPCCRRKRG---NLKKRR 297

Qy  301  EKEKTNKETETESGNENSGYNSDEQKTTDTASLP 334
         ::| |:|  | ::                   |||
Db  298  QTKKLRQKVEMKT--PATIQMNKRPQTPLLSLP 328
```

PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-0542217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS92356.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID NO 58528; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostic amino acid sequences of the invention.
CC   ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   463 AA;

Query Match   63.18;   Score 1318;   DB 22;   Length 463;
Best Local Similarity   99.6%;
Pred. No. 7.6e-101;
Matches 242;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

```
Qy       49 MWALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSR 108
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MWALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSR 60

Qy      109 LHGSAYLIVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYYF 168
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 LHGSAYLIVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYF 120

Qy      169 VPEPSDLQSAVSIIALTPQSNGTLTCVATWKSLKARKSAATVNLTVIRCPQDTGGGINIPG 228
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 VPEPSDLQSAVSIIALTPQSNGTLTCVATWKSLKARKSAATVNLTVIRCPQDTGGGININPG 180

Qy      229 VLSSLPSLGFSLPTWGKVGLGLAGTMLLLTPTCTLTIRCCCRRRCCGCNCCRCCCFCCRR 288
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 VLSSLPSLGFSLPTWGKVGLGLAGTMLLLTPTCTLTIRCCCRRRCCGCNCCRCCCFCCRR 240

Qy      289 KRG 291
            |||
Db      241 KRG 243
```

AAU75544
ID   AAU75544 standard; Protein; 370 AA.
XX
AC   AAU75544;
XX
DT   23-APR-2002 (first entry)
XX
DE   Mouse B7-like protein, B7-L_m1.
XX
KW   Mouse; B7-like protein; B7-Lj; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antithyroid; antiulcer; antiallergic;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Mus musculus.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13032.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 5; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I)
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions (pemphigus and
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and

RESULT 6

CC (pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC lymphoproliferative disorders such as multiple myeloma. The present
CC sequence represents the amino acid sequence of mouse B7-L_m1.
XX
SQ Sequence 370 AA;

Query Match 43.8%; Score 913.5; DB 23; Length 370;
Best Local Similarity 50.7%; Pred. No. 1.9e-67;
Matches 192; Conservative 50; Mismatches 102; Indels 35; Gaps 5;

```
Qy   14 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
         ::|||||  |:| ||||  ||||| ||: ||||||||||| ||: ||||| |||| ||
Db   21 SGSSYQIIEGPQNVTVLKDSEAHFNCTVTHGWKLLMWTLNQMVVLSLTTQGPIITNNRFT 80

Qy   74 SQRYDQGGNFTSEMIIHNVEPSDSDGNIRCSLQNSRLHGSAYLIYQVMGELFIPSVNLVVA 133
         |||||||||| || :|| ||||||||| ||||||| ||||| || | ||||||| |||
Db   81 YASYNSTDSFISELIIHDVQPSDSGSGVQCSLQNSHGFGSAFLSVQVMGTLNIPSNNLIVT 140

Qy  134 ENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLT 193
         ||||| ||||||| | ||||| |||| |||| ||||||
Db  141 EGEPCNVTCYAVGWTISLPDISWELEVPVSHSSYNSFLEPGCNFMRVLSVLDLTPLGNGTLT 200

Qy  194 CVATWKSLKARKSATVNLTVIRCPQDTGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGT 253
         |:| ||||||:|||||||||||  ||| |||
Db  201 CVAELKDLQASKLTVNLTVVQPPPD---------------SIGEGPALPTWAIILLAVAFS 248

Qy  254 MLITPTCLTIPIRCCCRRRCCGCNCCCRCCFCCRRKRGGFRIQFQ---KKSEKEKTNK--- 307
         ||:|:  |||||                    |:| |:||| |: ||
Db  249 LILILIIVLLIFCCC------------------------CASRREKEESTYQNEIRKSANMRTNKADP 293

Qy  308 ETETESGNENSGYNSDEQKTTDTASLPPKSCESSDPEQRNSSCGPFHQRADQRPRPASH 367
         |:|:|||||:||||||::|||||:||||||
Db  294 ETKLKSGKENYGYSSDEAKAAQTASLPPKSAEVSLPEKRSSSL--PYQELNKHQPGPATH 351

Qy  368 PQASFNLASPEKVSNTTVV 386
         | |:||||:||:|:||
Db  352 PRVSFDIASPQKVRNVTLV 370
```

RESULT 7

ID   AAU75547 standard; Protein; 631 AA.
XX
AC   AAU75547;
XX
DT   23-APR-2002 (first entry)
XX
DE   Rat B7-like protein, B7-1.
XX
OS   Ratus rattus.
XX
KW   Rat; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; antipruritic;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.

PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis
XX
PS   Disclosure; Fig 8; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of rat B7-1.
XX
SQ   Sequence 631 AA;

Query Match 27.4%; Score 572.5; DB 23; Length 631;
Best Local Similarity 27.8%; Pred. No. 7.1e-39;
Matches 163; Conservative 44; Mismatches 113; Indels 267; Gaps 9;

```
Qy   14 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT- 72
         ::| :|| |  :||| ||||  ||||| ||||||| ||:| ||||| |||| |
Db   69 SGIVGQVSKSVRNVTVLKDSEAHFNCTVTHGWKLLMWTLNQMVVLSLTTQGPIITNNRFE 128

Qy   73 ------------------------------------TSQRYDQGGNFTS 85
                                               |||||||||||
Db  129 KALLSCDYKFCSEEQSIRRIYWQKHDKWVLSVISGVPEVWPRYKNRTYASYNSTDSFIS 188

Qy   86 EMIIHNVEPSDSDGNIRCSLQNSRLHGSAYL---------------------- 115
         ||:|| |||||| ||:||||||:| |||:
Db  189 ELIIHDVQPSDSGSGVQCSLQNSHGFGSAFLSVQVDIANNYSFSLLGLIILSDRGTYTCVV 248

Qy  116 ----------TVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTWLPDISWELGLLV 161
                   ||||:||||||| ||:|
Db  249 QRYEGGSYVVKHLTTVEVMGTLNIPSNNLIVTEGEPCNVTCYAVGWTISLPDISWELEVPV 308

Qy  162 SHS-------------------------------------------SYFVPEPS 173
         |||                                          |:|:|||
Db  309 SHSLSVRADFPIPNITEYGNPSADIKRITCPASGGEPKPRLSWLENGRELNSYNSFLEPG 368
```

```
Qy     174  DLQSAVSILALTPQSNGTLLTCVATWKSLKARKSATVNL------------------- 211
                 :|:|| ||| |||| ||||||||||
Db     369  NFMRVLSVLDLPLGNGTLLTCVAELKDLQASKLTVNLGINTIISQDPESELYTISSQLD 428

Qy     212  -----------------TVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVG 247
                             |: ||||| |||||||:
Db     429  FNATYDHFIDCFIEYGDAHVSQNFIVVQPPPD----------SIGEGPALPTWAIIL 476

Qy     248  LGLAGTMLL------------------------------- 256
            |:||  :||:|
Db     477  LAVAFSLILIIIVLIIIIITWVKPEDPDEKQTVPFAWAGPDAVKAIIIFFIAITVIAV 536

Qy     257  TPTCLTLTIRCCCCRRRCCGCNCCCRCCFCCRRKRGFRIQFQ--KKSEKEKTNK----ETETE 312
            | :|| |||||
Db     537  IAIAIIIFCCCCASR----------REKEESTYQNEIRKSANMRTNKADPETKLK 582

Qy     313  SGNENSGYNSDEQKTTDTASLPPKSCESSDPEQRNSSCGPPHQRADQ 359
            |:|:||| |||||  |:|  |:|
Db     583  SGKENYGYSSDECITVKFRRCFPRRRNEASRETNKNLYIGPVEAAAEQ 629
```

RESULT 8

AAU75545

ID   AAU75545 standard; Protein; 270 AA.
XX
AC   AAU75545;
XX
DT   23-APR-2002 (first entry)
XX
DE   Mouse B7-like protein, B7-L_m2.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Mus musculus.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13033.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
PS   Claim 13; Fig 6; 135pp; English.
XX

CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of mouse B7-L_m2.
XX
SQ   Sequence 270 AA;

Query Match   27.1%; Score 566.5; DB 23; Length 270;
Best Local Similarity   34.6%;
Matches 131; Conservative   42; Mismatches 71; Indels 135; Gaps 5;

```
Qy      14  SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIIINDRFT 73
            |||| :||||||| |:|  |:|| |||||| |||:||| |:||||:||  ||||||||||
Db      21  SGSSYQIIEGPQNVTVLKDSEAHFNCTVTHGWKLLMWTLNQMVVLSLTTQPIITNNRFT 80

Qy      74  SQRYDQGGNFTSEMIIHNVEPSDSDGNTRCSLQNSRLHGSAYLTVQVWGELFIPSVNLVVA 133
            | :| ::|:| :|||||:|:|||| | :|||||:|  ||:|:|:|:|
Db      81  YASYNSTDSFISELIIHDVQPSDSGSSVQCSLQNSHGFGSAFLSVQ---------------125

Qy     134  ENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLT 193
Db     126  ------------------------------------------------------------125

Qy     194  CVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGT 253
                                                  :||||  | :|||||||||: :|
Db     126  --------------------------------------DSIGEGPALPTWAIILAVAFS 148

Qy     254  MLIIPTCTLTIRCCCCRRRCCGCNCCCRCCFCCRRKRGFRIQFQ---KKSEKEKTNK--- 307
            :|  ||| |:|                          ||  :||:  ::|:: |||
Db     149  ----------------------------------CASRREKEESTYQNEIRKSANMRTNKADP 193

Qy     308  ETETESGNENSGYNSDEQKTTDTASLPPKSCESSDPEQRNSSCGPPHQRADQRPRPASH 367
            |: :||  ||:|  ||:||||||||:| ||||| |:| |:|
Db     194  ETKLKSGKENYGYSSDEAKAAQTASLPPKSAEVSLPERSSSL--PYQELNKHQPGPATH 251

Qy     368  PQASENLASPEKVSNTTVV 386
            | :|  :|||:||:|  :|
Db     252  PRVSEDIASPQKVRNVTLV 270
```

RESULT 9

AAU75546

ID   AAU75546 standard; Protein; 223 AA.
XX
AC   AAU75546;
XX
DT   23-APR-2002 (first entry)
XX
DE   Mouse B7-like protein, B7-L_m3.
XX

KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic;
KW   antiinflammatory; dermatological; antipsoriatic; antirheumatic;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Mus musculus.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13034.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 7; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disorders (e.g.
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions, and for
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of mouse B7-L_m3.
XX
SQ   Sequence   223 AA;

Query Match   223 AA;   23.28;   Score 485;   DB 23;   Length 223;

```
Best Local Similarity  31.4%;  Pred. No. 3.4e-32;
Matches  118;  Conservative  37;  Mismatches  45;  Indels  176;  Gaps  6;

Qy   14 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIIINDRFT 73
         :||: |||||||:||||:||||||:|::|| |||::||  :  |||||||:: ||||:||
Db   21 SGSSYQIIEGPQNVVLKDSEAHFNCTVTHGWKLLMWTLNQMVVLSLTTQGPIITNNRFT 80

Qy   74 SQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVVA 133
         : || ||||:||:|| |: |:||||||::|:|||||||:|||||||||||||||||||
Db   81 YASYNSTDSFISELIIHDVQPSDSGSVQCSLQNSHGFGSAFLSVQ----------------- 125

Qy  134 ENEPCEVTCLPSHWTWLPDISWELGLLIVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLT 193
         --------------------------:|:-----------:|:--------------------
Db  126 --------------------------ESTY-------QNEI-------------------- 133

Qy  194 CVATWKSLKARKSATVNLIVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGT 253
         :|| |||| |:||||||:||||||||| | ||:||||| |||||:||:|||||  ||||
Db  134 ----------RKSANM----------------------------------------- 139

Qy  254 MLLTPICTLIIRCCCCRRRCCGCNCCRRCCGGCNCCCRCCFCCRRKRGFRIQFOKKSEKEKTNK---ETE 310
                                                                     :|||
Db  140 -------------------------------------------------------RTNKADPETK 149

Qy  311 TESGNENSGYNSSDEQKTTDFASLPKSCESSDPEQRNSSCGPPHQRADORPPRRASHPQA 370
         :|:|                      ||:|| |:     |  :|:     |:|:
Db  150 LKSGKENYGYSSDEAKAAQTASLPPKSAEVSLPERSSSSL--PYQELNKHQPGPATHPRV 207

Qy  371 SFNLASPEKYSNTTVV 386
         ||:||||||:||| | ||
Db  208 SFDIASPQKVRNVTLV 223
```

RESULT 10
AAB81925

ID   AAB81925 standard; protein; 404 AA.
XX
AC   AAB81925;
XX
DT   15-JUN-2001 (first entry)
XX
DE   Extracorporeal circulation material receptor protein.
XX
KW   Extracorporeal circulation; carbonyl stress product; receptor;
KW   diabetes; vascular lesion; excretory dysfunction.
XX
OS   Unidentified.
XX
PN   WO200118060-A1.
XX
PD   15-MAR-2001.
XX
PF   08-SEP-2000; 2000WO-JP06172.
XX
PR   08-SEP-1999;   99JP-0254463.
XX
PA   (TORA) TORAY IND INC.
XX
PI   Shimizu S, Kubota M, Akiyama H, Usui M;
XX
DR   WPI; 2001-290314/30.
XX
PS   Claim 1; Page 31-32; 36pp; Japanese.
XX
PT   Material for extracorporeal circulation, applicable in selective
PT   elimination of diabetic complication factors such as carbonyl stress
PT   products caused by abnormally promoted carbonyl stress from excretory
PT   dysfunction in vascular lesions -
XX
CC   The present invention describes a material for extracorporeal circulation
CC   which is made from a water-insoluble carrier immobilized with a protein
CC   having the sequence shown here. The materials of the invention, including
CC   adsorbents, are for extracorporeal circulation, which are applicable in
CC   the selective elimination of diabetic complication factors from a body

CC fluid, and are therefore useful in treating vascular lesions like
CC arteriosclerosis due to carbonyl stress products caused by abnormally
CC promoted carbonyl stress from excretory dysfunction.
XX
SQ Sequence 404 AA;

Query Match   7.0%; Score 145.5; DB 22; Length 404;
Best Local Similarity 23.5%; Pred. No. 0.001;
Matches   77; Conservative 34; Mismatches 94; Indels 123; Gaps 15;

```
Qy   64 EPIITNDRFTS-----QRYDQGGNFT--SEMIHNVEPSDSGNIR----CSLQNSRLHGS 112
           ||:||  |||     |:|:|                            |:||  ||||||
Db  162 KPLVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHR 218

Qy  113 AYLIVQVMGELFIP------SVNLVVAENEP----------CEVTCLPSHWTWLPDIS 154
                                                     |||  |   |  | ||
Db  219 ALRTAPIQPRVWEPVPLEEVQLVV--EPEGGAVAPGGTVTLTCEVPAQPS----PQIH 270

Qy  155 WELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVI 214
                                               ||||  |     |:|  :|:|
Db  271 WMKD-------GVPLPLPPSPVLILPEIGPDQDGTYSCVATHSSHGPQESRAVSISII 321

Qy  215 RCPQDTGGGINIPGVLSLSPLSLGFSLPTWGKVGLGLAGTMLLT------PTCTLTIRCC 267
           |                                  ||| ||       |:||||||
Db  322 E-PGEEG-----------PTAGSVGGSGLGTLALALGILGGLGTAALLIGVI 361

Qy  268 CCRRRCCGCNCCRCCFCCRRKRRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQ 325
                                   :||  |:| ||     |  | |:||
Db  362 LWQRR-----------QRRGEERKAPENQEEEERAELN------- 389

Qy  326 KTTDTASLPPKSCESSDPEQRNSSCGPP 353
             :|||      |     |  ||:||
Db  390 -----------QSEEPEAGESSTGGP 404
```

RESULT 11
AAE23219
ID   AAE23219 standard; Protein; 404 AA.
XX
AC   AAE23219;
XX
DT   27-AUG-2002 (first entry)
XX
DE   Human receptor for advanced glycosylation end product (RAGE) protein.
XX
KW   Human; Receptor for advanced glycosylation end product; RAGE; cardiant;
KW   tissue growth; neointimal formation; blood vessel; restenosis; diabetes;
KW   myocardial infarction; angioplasty; peripheral vascular surgery; angina;
KW   transgenic animal; acute thrombotic stroke; venous thrombosis.
OS   Homo sapiens.
XX
PN   WO20230889-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US32036.
XX
PR   13-OCT-2000; 2000US-0687528.
XX
PA   (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI   Stern DM, Schmidt A, Marso S, Topol E, Lincoff AM;
XX
DR   WPI; 2002-426260/45.
DR   N-PSDB; AAD36952.
XX
PT   Inhibiting new tissue growth or neointimal formation in blood vessels
PT   of subject suffering from diabetes, stroke and preventing restenosis,
PT   comprises administering inhibitor of receptor for advance glycation end
PT   product -
XX
PS   Disclosure; Page 16; 43pp; English.

XX
CC   The invention relates to a method for inhibiting new tissue growth or
CC   neointimal formation in blood vessels in a subject that has experienced
CC   blood vessel injury and preventing exaggerated restenosis in a diabetic
CC   subject. The method comprises administering an inhibitor of receptor for
CC   advanced glycation/glycosylation end product (RAGE), so as to inhibit new
CC   tissue growth or neointimal formation in subject's blood vessels and
CC   preventing restenosis in the subject. The method is useful for inhibiting
CC   new tissue growth or neointimal formation in blood vessels in a subject
CC   like non-human animal, a transgenic non-human animal or a human suffering
CC   from diabetes, myocardial infarction, abrupt closure following angioplasty or
CC   angina, myocardial infarction, acute thrombotic stroke, venous thrombosis, unstable
CC   stent placement, or thrombosis as a result of peripheral vascular surgery
CC   The method is also useful for preventing restenosis and for determining
CC   whether a compound inhibits new tissue growth in a blood vessel in a
CC   subject. The present sequence is human receptor for advanced
CC   glycosylation end product (RAGE) protein.
XX
SQ Sequence 404 AA;

Query Match   7.08%; Score 145.5; DB 23; Length 404;
Best Local Similarity 23.5%; Pred. No. 0.001;
Matches   77; Conservative 34; Mismatches 94; Indels 123; Gaps 15;

```
Qy   64 EPIITNDRFTS-----QRYDQGGNFT--SEMIHNVEPSDSGNIR----CSLQNSRLHGS 112
           ||:||  |||     |:|:|                            |:||  ||||||
Db  162 KPLVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHR 218

Qy  113 AYLIVQVMGELFIP------SVNLVVAENEP----------CEVTCLPSHWTWLPDIS 154
                                                     |||  |   |  | ||
Db  219 ALRTAPIQPRVWEPVPLEEVQLVV--EPEGGAVAPGGTVTLTCEVPAQPS----PQIH 270

Qy  155 WELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVI 214
                                               ||||  |     |:|  :|:|
Db  271 WMKD-------GVPLPLPPSPVLILPEIGPDQDGTYSCVATHSSHGPQESRAVSISII 321

Qy  215 RCPQDTGGGINIPGVLSLSPLSLGFSLPTWGKVGLGLAGTMLLT------PTCTLTIRCC 267
           |                                  ||| ||       |:||||||
Db  322 E-PGEEG-----------PTAGSVGGSGLGTLALALGILGGLGTAALLIGVI 361

Qy  268 CCRRRCCGCNCCRCCFCCRRKRRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQ 325
                                   :||  |:| ||     |  | |:||
Db  362 LWQRR-----------QRRGEERKAPENQEEEERAELN------- 389

Qy  326 KTTDTASLPPKSCESSDPEQRNSSCGPP 353
             :|||      |     |  ||:||
Db  390 -----------QSEEPEAGESSTGGP 404
```

RESULT 12
AAU77543
ID   AAU77543 standard; Protein; 404 AA.
XX
AC   AAU77543;
XX
DT   05-JUN-2002 (first entry)
XX
DE   Human receptor for advanced glycosylation end product (RAGE).
XX
KW   Receptor for advanced glycation end product; RAGE; receptor;
KW   amyloid beta peptide; blood-brain barrier; neurovascular stress;
KW   cerebral vasoconstriction suppressor; cerebral blood flow enhancer;
KW   cerebral amyloid angiopathy; transgenic animal; amyloid beta precursor;
KW   Alzheimer's disease; Down's syndrome; head trauma; stroke; human.
OS   Homo sapiens.
XX
PN   WO20021451519-A1.
XX
PD   21-FEB-2002.
XX
PF   14-AUG-2001; 2001WO-US25416.
XX

Left column (RESULT 13)

PR   14-AUG-2000; 2000US-0638648.
XX
PA   (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI   Stern DM, Schmidt AM, Yan SD, Zlokovic B;
XX
DR   WPI; 2002-257610/30.
DR   N-PSDB; ABK10856.
XX

PT   Ameliorating neurovascular stress and decreasing cerebral
PT   vasoconstriction in subject suffering from chronic/acute cerebral
PT   amyloid angiopathy, by administering inhibitor of receptor for advanced
PT   glycation endproduct.
XX
PS   Disclosure; Page 16; 68pp; English.
XX

CC   The invention describes a method of ameliorating neurovascular stress,
CC   and decreasing cerebral vasoconstriction in subject suffering from
CC   chronic or acute cerebral amyloid angiopathy, comprising administering
CC   an inhibitor (I) of receptor for advanced glycation end product (RAGE).
CC   (I) inhibits transcytosis of amyloid beta peptides across blood-brain
CC   barrier, thus decreasing cerebral vasoconstriction and increasing
CC   cerebral blood flow. (I) is useful for treating amyloid angiopathy in a
CC   subject, decreasing cerebral vasoconstriction in a transgenic non-human
CC   animal (preferably, transgenic mouse overexpressing mutant human amyloid
CC   beta precursor protein) or a human, suffering from chronic or acute
CC   cerebral amyloid angiopathy, preferably, Alzheimer's disease, and for
CC   ameliorating neurovascular stress comprising cerebral amyloid angiopathy
CC   in a subject, where the neurovascular stress is caused by Alzheimer's
CC   disease, aging, Down's syndrome, head trauma or stroke. This is the
CC   amino acid sequence of human receptor for advanced glycation end
CC   product (RAGE) described in the invention.
XX
SQ   Sequence   404 AA;

Query Match   7.0%;
Best Local Similarity   23.5%;
Matches 77; Conservative 34; Mismatches 94; Indels 123; Gaps 15;
                     Score 145.5; DB 23; Length 404;

```
Qy   64 EPIITNDRFTS----QRYDQGGNFT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGS 112
          :|:| |:|          || |:|          |||:  :|::::|
Db  162 KPLVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGDPRPTFSCSFSPGLPRHR 218

Qy  113 AYLTVQVWGELFIP----SVNLVVAENEP-----------CEVTCLPSHWTWLPDIS 154
Db  219 ALRTAPIQPRWEPVPLEEVQLVV--EPEGGAVAPGGTVTLTCEVPAQPS----PQIH 270

Qy  155 WELGLLVSHSSYYFVPEPSDLQSAVSIALTPQSNGTLTCVATWKSLKARKSATVNLTVI 214
Db  271 WMKD----GVPLPLPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISII 321

Qy  215 RCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLT------PTCTLTIRCC 267
Db  322 E-PGEEG--------------------------PTAGSVGGSGLGTLALALGILGGLGTRAALLIGVI 361

Qy  268 CCRRRCCGCNCCCRCCFCCRRKRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQ 325
Db  362 LWQRR-----------------QRRGEERKAPENQEEEERAELN-------389

Qy  326 KTTDTASLPPKSCESSDPEQRNSSCGPP 353
Db  390 ---------------QSEEPEAGESSTGGP 404
```

RESULT 13
AAM48745
ID   AAM48745 standard; protein; 404 AA.
XX
AC   AAM48745;
XX
DT   02-APR-2002 (first entry)
XX
DE   Human RAGE protein SEQ ID NO 1.

Right column (RESULT 14)

XX
KW   Human; RAGE; receptor for advanced glycated endproduct; receptor;
KW   antidiabetic; neuroprotective; cytostatic; antiinflammatory; receptor;
KW   nephrotropic; dermatological; antiarteriosclerotic; nootropic; vasotropic;
KW   Alzheimer's disease; cancer; inflammation; kidney failure; diabetes;
KW   systemic lupus; nephritis; erectile dysfunction; atherosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO200192892-A2.
XX
DR   WPI; 2002-114372/15.
PI   Shahbaz M;
XX
PA   (TRAN-) TRANS TECH PHARMA.
XX
PR   05-MAR-2001; 2001US-0799152.
PR   30-MAY-2000; 2000US-207342P.
XX
PD   30-MAY-2001; 2001WO-US17447.
XX
PD   06-DEC-2001.
XX

PT   Detecting a receptor for advanced glycated endproducts (RAGE)
PT   modulators, for treating e.g. cancer, diabetes or inflammation,
PT   comprises measuring the amount of bound anti-RAGE antibody
XX
PS   Claim 1; Fig 2; 49pp; English.
XX

CC   The invention relates to detecting receptor for advanced glycated
CC   endproducts (RAGE) modulators comprises determining the amount of RAGE
CC   protein or its fragment bound to the pre-adsorbed ligand by measuring the
CC   amount of anti-RAGE antibody bound to the solid surface. The method is
CC   useful for rapid, high-throughput identification of compounds that
CC   modulate RAGE. The compounds are useful for treating symptoms of diabetes
CC   and symptoms of diabetic late complications, amyloidoses, Alzheimer's
CC   disease, cancer, inflammation, kidney failure, systemic lupus nephritis
CC   or inflammatory lupus nephritis, erectile dysfunction and
CC   atherosclerosis.
XX
SQ   Sequence   404 AA;

Query Match   7.0%;
Best Local Similarity   23.5%;
Matches 77; Conservative 34; Mismatches 94; Indels 123; Gaps 15;
                     Score 145.5; DB 23; Length 404;

```
Qy   64 EPIITNDRFTS----QRYDQGGNFT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGS 112
          :|:| |:|          || |:|          |||:  :|::::|
Db  162 KPLVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGDPRPTFSCSFSPGLPRHR 218

Qy  113 AYLTVQVWGELFIP----SVNLVVAENEP-----------CEVTCLPSHWTWLPDIS 154
Db  219 ALRTAPIQPRWEPVPLEEVQLVV--EPEGGAVAPGGTVTLTCEVPAQPS----PQIH 270

Qy  155 WELGLLVSHSSYYFVPEPSDLQSAVSIALTPQSNGTLTCVATWKSLKARKSATVNLTVI 214
Db  271 WMKD----GVPLPLPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISII 321

Qy  215 RCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLT------PTCTLTIRCC 267
Db  322 E-PGEEG--------------------------PTAGSVGGSGLGTLALALGILGGLGTRAALLIGVI 361

Qy  268 CCRRRCCGCNCCCRCCFCCRRKRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQ 325
Db  362 LWQRR-----------------QRRGEERKAPENQEEEERAELN-------389

Qy  326 KTTDTASLPPKSCESSDPEQRNSSCGPP 353
Db  390 ---------------QSEEPEAGESSTGGP 404
```

RESULT 14

AAW81030
ID   AAW81030 standard; Protein; 1496 AA.
XX
AC   AAW81030;
XX
DT   10-MAY-1999 (first entry)
XX
DE   Melanoma associated antigen MG50.
XX
OS   Homo sapiens.
XX
KW   MG50; melanoma gene-50; melanoma associated antigen; human;
KW   T cell epitope; cancer; lung cancer; rhabdomyosarcoma; diagnosis;
KW   therapy; vaccine.
XX
PN   WO9855133-A1.
XX
PD   10-DEC-1998.
XX
PF   04-JUN-1998; 98WO-US11533.
XX
PR   06-JUN-1997; 97US-0870941.
XX
PA   (REGC) UNIV CALIFORNIA.
PA   (UYSC-) UNIV SOUTHERN CALIFORNIA.
XX
PI   Deans RJ, Kan-Mitchell J, Minev BR, Mitchell MS;
XX
DR   WPI; 1999-080820/07.
DR   N-PSDB; AAV99922.
XX
PT   New MG50 melanoma associated antigen fragments - used to develop
PT   products for the detection, treatment and prevention of
PT   MG50-expressing cancers, e.g. melanoma, lung cancer or
PT   rhabdomyosarcoma.
XX
PS   Claim 1; Page 45-49; 79pp; English.
XX
CC   This polypeptide comprises a portion of a new human melanoma
CC   associated antigen, designated MG50. The amino acid sequence was
CC   deduced from a cDNA clone (see AAV99922) isolated from melanoma
CC   cell line MSM M-1 cDNA by subtractive hybridisation. The 5' region
CC   of MG50 cDNA was not obtained. MG50 mRNA has been detected in
CC   melanoma, lung carcinoma and rhabdomyosarcoma cells, foetal brain,
CC   foetal heart and human placenta. The invention also provides T
CC   helper T cell epitopes (see AAW81031-54) from MG50, including
CC   cell epitopes (see AAW81031-54) from MG50, including cytotoxic and
CC   or an MG50 T cell epitope, recombinant vectors, and antigen
CC   presenting cells. Methods are provided for identifying an MG50
CC   melanoma associated antigen in an individual and for identifying an
CC   immune response against an MG50 melanoma associated antigen, as
CC   well as methods of stimulating T lymphocytes that are reactive
CC   against cancer cells expressing MG50 and for treating an individual
CC   having cancer cells that express MG50. The products and methods
CC   can be used for the detection, treatment and prevention of
CC   MG50-expressing cancers, e.g. melanomas, lung cancer or
CC   rhabdomyosarcoma.
XX
SQ   Sequence 1496 AA;

Query Match   25.1%; Score 138; DB 20; Length 1496;
Best Local Similarity   6.6%;
Matches   78; Conservative   38; Mismatches 127; Indels   68; Gaps 17;

Qy   2 VAGAMENRDPP----GSGSGNEVIEGPQNARVLKGSQARFNCTVS-QGMKLIMWALSDMV 56
          ||| |: :| ||| || | || | : | | |||
Db 339 VAGEVKTQEVTLRYFGSPARPTEVIQPQNTEVLVGESVTLECSATGHPPRISWTRGDRT 398

Qy   57 VLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIHNVEPSDSGNIRCSLONS--RLHGSAY 114
           |: :| ||| | | | |||: |: ||| | |: |: |:
Db 399 PLPVDPRVNITPS----------GG-----LYIQNVVQGDSGEYACSATNNIDSVHATAF 443

Qy 115 LTVQVWGELFIPSVNLVVAENE-----PCEVTCLPSHWTWLPDISW-ELGLLVSHSSYPV 169

Db 444 IIVQALPQFTVTPQDRVVIEGQTVFDFQCEAKGNPP----PVIAWTKGGSQLSVDRHILV 498
           ||| :| | :||| ||:| || | ||| | :| | |
Qy 170 PEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-------IRCPQDT- 220
           || : | || ||| |:|| | || ||||| | |||
Db 499 LSSGTLR--ISGVALHDQ--GQYECQAV--NIIGSQKVVAHLTVQPRVTPVFASIPSDTT 552

Qy 221 --GGGINIPGVLSSLPSLGFSLP--TWGSKVGLGLL-AGTMLLTPTCTLTI---------- 264
          :| || | | | | || | | | | |||||:|| | ||
Db 553 VEVGANVQLP-----CSSQGEPEPAITWNKDGVGVTESGSKFHISPEGFLTINDVGPADAG 607

Qy 265 RCCCCRRCCG 275
          || | | ||
Db 608 RYECVARNTIG 618

RESULT 15
AAY70469
ID   AAY70469 standard; Protein; 1496 AA.
XX
AC   AAY70469;
XX
DT   21-JUN-2000 (first entry)
XX
DE   Human p53 target molecule, PRG2 protein.
XX
OS   Homo sapiens.
XX
KW   PRG2; p53 target; human; modulate; cell proliferation; immunomodulatory;
KW   chromosome 2p24.3; cytostatic; gene therapy; tumour cell; inducer;
KW   diagnosis; therapeutic; proliferative disease; cell cycle arrest; cancer;
KW   treatment; apoptosis; knockout animal; cancer susceptibility; dPxn; hPxn;
KW   peroxidasin; redox-regulation; reactive oxygen species; ROS.
XX
FH   Key   Location/Qualifiers
FH
FT   Peptide   1..29
FT   /label= Signal_peptide
FT   Peptide   30..1496
FT   /label= Mature_human_PRG2_protein
FT   /note= "Homologue of Drosophila peroxidasin, dPxn"
XX
PN   WO200012526-A1.
XX
PD   09-MAR-2000.
XX
PF   27-AUG-1999; 99WO-US19551.
XX
PR   28-AUG-1998; 98US-0098251.
XX
PA   (UYPR-) UNIV PRINCETON.
XX
PI   Horikoshi N, Shenk T;
XX
DR   WPI; 2000-246724/21.
DR   N-PSDB; AAZ51671.
XX
PT   New p53-inducible isolated nucleic acid molecule including open reading
PT   frame encoding human homolog of Drosophila melanogaster peroxidasin,
PT   useful e.g. in detection and treatment of cancer
XX
PS   Disclosure; Page 71-74; 83pp; English.
XX
CC   The present sequence is the protein encoded by PRG2 gene, that is
CC   upregulated in response to induction of p53 activity in human colon
CC   cancer EB1 cells. This sequence is the human homologue (hPxn) of
CC   Drosophila peroxidasin gene dPxn, that is expressed in heart, placenta,
CC   spleen, ovary and intestines. PRG2 is involved in p53-mediated growth
CC   suppression pathways and plays a role in redox regulation. It is a
CC   haem-peroxidase that increases the intracellular content of reactive
CC   oxygen species (ROS). They are potential targets of p53 regulatory
CC   activity and are useful for modulation of cellular proliferation. PRG2
CC   gene is localised to human chromosome 2p24.3. The PRG target molecules
CC   have cytostatic and immunomodulatory activity. PRG polynucleotides,

```
CC   proteins and antibodies are useful as diagnostic and therapeutic agents
CC   for detection and treatment of cancer and other proliferative diseases.
CC   The gene/cDNA may be used for gene therapy, to restore a gene function
CC   downstream of p53, that cannot be activated in the p53-deficient tumour
CC   cell. Antibodies can be used as inducers of cell cycle arrest and/or
CC   apoptosis. The DNA sequences can be used to generate 'knockout' animals
CC   as a model of cancer susceptibility.
XX
SQ   Sequence   1496 AA;

     Query Match    6.6%;   Score 138;   DB 21;   Length 1496;
     Best Local Similarity   25.1%;   Pred. No. 0.022;
     Matches   78;   Conservative   38;   Mismatches 127;   Indels   68;   Gaps   17;

Qy      2  VAGAMENRDPP----GSGSGNEVIEGPQNARVLKGSQARFNCTVS-QGWKLIMWALSDMV   56
            ||  |  :  ::      |    ||  :: || |  |:  : :      |||| |:
Db    339  VAGEVKTQEVTLRYFGSPARPTFVIQPQNTEVLVGESVTLECSATGHPPPRISWTRGDRT  398

Qy     57  VLSVRPMEPIIINDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNS--RLHGSAY  114
            | | |   || |   : | |   |   |    ||   :|||  ::  |:      :|::
Db    399  PLPVDPRVNITPS--------GG----LYIQNVVQGDSGEYACSATNNIDSVHATAF   443

Qy    115  LTYQVMGELFIPSVNLVVAENE-----PCEVTCLPSHWTWLPDISW-ELGLLVSHSSYFV  169
            :  |: || :  : : |   :         :   |:  ::  : :|  :: | :
Db    444  IIVQALPQFTVTPQDRVVIEGQTVDFQCEAKGNPP------PVIAWTKGGSQLSVDRRHLV  498

Qy    170  PEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV----------IRCPQDT-  220
            |: | :| : ||| |  |:| |  | |  :     | |  :|            | | :
Db    499  LSSGTLR--ISGVALHDQ--GQYECQAV--NIIGSQKVVAHLTVQPRVTPVFASIPSDTT  552

Qy    221  ---GGGINIPGVLSSLPSLGFSLP--TWGKIVGLGL--AGTMLLTPTCTLIT-------  264
             :  ||| | || | | |   :||    ||  |   |||| || |  ||
Db    553  VEVGANVQLP-------CSSQGPEPAITWNKLDGVQVTESGKFHISPEGFLTINDVGPADAG  607

Qy    265  RCCCCRRRCCG  275
            |  |
Db    608  RYECVARNTIG  618
```

```
              GenCore version 5.1.4_p5_4578
              Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:        April 28, 2003, 16:24:32 ; Search time 37.1283 Seconds
                                           (without alignments)
                                           2142.147 Million cell updates/sec

Title:         2077
Perfect score: 2077
Sequence:      1 MERHLLTVPEAVGSGSGNEV.........HPQASFNLASPEKVSNTTVV 386
Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5
Searched:      671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters:        671580
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     1: sp_archea:*
               2: sp_bacteria:*
               3: sp_fungi:*
               4: sp_human:*
               5: sp_invertebrate:*
               6: sp_mammal:*
               7: sp_mhc:*
               8: sp_organelle:*
               9: sp_phage:*
              10: sp_plant:*
              11: sp_rodent:*
              12: sp_virus:*
              13: sp_vertebrate:*
              14: sp_unclassified:*
              15: sp_rvirus:*
              16: sp_bacterial:*
              17: sp_archeap:*
              SPTREMBL_21:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1694	81.6	315	4	Q9NSI5	Q9nsi5 homo sapien
2	579.5	27.9	270	11	Q9D8G2	q9d8g2 mus musculu
3	142.5	6.9	344	13	Q9DF61	q9df61 gallus gall
4	141	6.8	168	11	Q9D9I2	q9d9i2 mus musculu
5	137.5	6.6	315	13	Q9DG15	q9dg15 gallus gall
6	137.5	6.6	344	13	Q9DGI5	q9dgi5 gallus gall
7	137.5	6.6	344	11	Q9PJ0	q9pj0 mus musculu
8	136.5	6.6	173	11	Q93242	q93242 gallus gall
9	135	6.5	1496	4	Q9D4K2	q9d4k2 mus musculu
10	133.5	6.4	344	4	Q92626	q92626 homo sapien
11	133	6.4	1427	13	Q9P121	q9p121 homo sapien
12	131	6.3	697	11	Q91562	q91562 xenopus lae
13	131	6.3	697	11	Q91562	q91562 homo sapien
14	129.5	6.2	164	5	Q9DBP0	q9dbp0 mus musculu
15	129	6.2	6632	5	Q9Z290	q9z290 mus musculu
16	128	6.2	1445	11	Q22048	q22048 caenorhabdi
					O01761	o01761 caenorhabdi
					Q63155	q63155 rattus norv
17	127.5	6.1	197	5	Q17641	Q17641 caenorhabdi
18	127.5	6.1	869	4	O15146	O15146 homo sapien
19	127.5	6.1	871	11	Q61987	Q61987 mus musculu
20	127.5	6.1	881	11	Q61988	Q61988 mus musculu
21	127	6.1	313	13	O57596	O57596 gallus gall
22	126	6.1	6620	4	Q96AA2	Q96aa2 homo sapien
23	125.5	6.1	166	5	Q95QY1	Q95qy1 caenorhabdi
24	122.5	5.9	188	5	O18238	O18238 caenorhabdi
25	122.5	5.9	455	4	Q9UIR0	Q9uir0 homo sapien
26	122.5	5.9	4370	4	Q9H3V5	Q9h3v5 homo sapien
27	120	5.8	695	11	Q9JJ09	Q9jj09 rattus norv
28	120	5.8	959	5	Q9N9Y9	Q9n9y9 drosophila
29	120	5.8	968	5	Q9W4T9	Q9w4t9 drosophila
30	120	5.8	975	5	O97174	O97174 drosophila
31	119.5	5.8	6632	5	O17362	O17362 caenorhabdi
32	117.5	5.7	390	4	Q9H1X9	Q9h1x9 homo sapien
33	117.5	5.7	5636	4	Q96RW7	Q96rw7 homo sapien
34	117	5.7	345	5	Q9W4U1	Q9w4u1 drosophila
35	116.5	5.6	275	11	Q9JK39	Q9jk39 mus musculu
36	116	5.6	2221	5	Q9U1M1	Q9u1m1 drosophila
37	115.5	5.6	1596	4	Q9HCL6	Q9hcl6 homo sapien
38	114.5	5.5	152	5	Q9XVX3	Q9xvx3 caenorhabdi
39	114	5.5	46	4	O16861	O16861 homo sapien
40	114	5.5	569	11	Q920G3	Q920g3 mus musculu
41	114	5.5	868	11	Q62838	Q62838 rattus norv
42	114	5.5	1395	5	O44924	O44924 drosophila
43	114	5.5	1395	5	Q9W213	Q9w213 drosophila
44	113.5	5.5	2673	4	Q96SC3	Q96sc3 homo sapien
45	113.5	5.5	196	4	Q9H0T1	Q9h0t1 homo sapien

ALIGNMENTS

RESULT 1
Q9NSI5

```
ID   Q9NSI5    PRELIMINARY;    PRT;   315 AA.
AC   Q9NSI5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGSF5 protein (Fragment).
GN   IGSF5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Park H.S.,
RA   Toyoda A., Ishii K., Totoki Y., Choi D.K., Soeda E., Ohki M.,
RA   Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.L.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL163280; CAB90447.1; -.
DR   Interpro; IPR000561; EGF-like.
DR   Interpro; IPR003599; Ig.
DR   Interpro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00409; IG; 2.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
FT   NON_TER        1      1
FT   NON_TER      315    315
SQ   SEQUENCE   315 AA;   34599 MW;  B69CB23570FA3D1 CRC64;
```

RESULT 2

```
Query Match
Best Local Similarity   99.7%;   81.6%;
Score 1694;  DB 4;  Length 315;
Matches 314;  Conservative 0;  Pred. No. 7; 3e-144;  Mismatches 1;  Indels 0;  Gaps 0;

Qy   14  SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT   73
         |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    1  SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT   60

Qy   74  SQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVVA  133
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  SQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVVA  120

Qy  134  ENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLT  193
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  ENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLT  180

Qy  194  CVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGT  253
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  CVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGT  240

Qy  254  MLLTPTCTLTIRCCCCRRRCCGCNCCCRCCCFCCRRKRGFRIQFQKKSEKEKTNKEIETES  313
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  MLLTPTCTLTIRCCCCRRRCCGCNCCCRCCCFCCRRKRGFRIQFQKKSEKEKTNKEIETES  300

Qy  314  GNENSGYNSDEQKTI  328
         ||||||||||||||
Db  301  GNENSGYNSDEQKTI  315
```

```
RESULT 2
ID   Q9D8G2        PRELIMINARY;      PRT;   270 AA.
AC   Q9D8G2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   2010003D20Rik protein (RIKEN cDNA 2010003D20 gene).
GN   2010003D20Rik.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Strausberg R.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK008060; BAB25436.1; -.
DR   EMBL; BC004806; AAH04806.1; -.
DR   MGI; MGI:1919308; 2010003D20Rik.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   SMART; SM00409; IG; 1.
DR   PROSITE; PS00290; IG_MHC; UNKNOWN_1.
SQ   SEQUENCE   270 AA;  29604 MW;  A39C273DAA950DFE CRC64;
```

RESULT 3

```
Query Match
Best Local Similarity   27.9%;   34.3%;
Score 579.5;  DB 11;  Length 270;
Matches 133;  Conservative 46;  Pred. No. 6; 3e-44;  Mismatches 74;  Indels 135;  Gaps 5;

Qy    5  LITVPEAVGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPME   64
            ::|  .|| .|| ||:||||||.||||.|.:.||||||.|||||::|.|.||||.::
Db   12  LVILAQLTASGSSYQIIEGPQNVTVLKDSEAHFNCTVTHGWKLLMMVTLNQMVVLSLTTQG  71

Qy   65  PIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELF  124
         |||||:||| :|:.|:|::| ||:.|.:.||||||:.|.|||||.|.|:|.||.||
Db   72  PIITNNRFTYASYNSTDSFISELIIHDVQPSDSGSVQCSLQNSHGFGSAFLSVQ------ 125

Qy  125  IPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILAL  184
            .:|      ..:      .:      :|      ||:|:|
Db  126  ---------------------------------DSIGEEGPALPTWA             139

Qy  185  TPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWG  244
         |:       ::::::      :|.::.:|||:
Db  126  CASREKEESTYQNEIRKSA                                         184

Qy  245  KVGLGLAGTMLLTPTCTLTIRCCCCRRRCCGCNCCCRCCCFCCRRKRGFRIQFQ---     301
            .:|:|    ::||::      :||
Db  185  NMRTNKADPETKLKSGKENYGYSSDEAKAAQTASLPPKSAEVSLPEKRSSSL-PYQELN 242

Qy  302  KEKTNK--EYETESGNENSGYNSDEQKTTETALSLPPKSCESSDPEQRNSSCGPPHQRAD  358
            ::|    :|:|:::      :|      :::      .:|||
Db  243  KHQPGPATHPRVSFDIASPQKVRNVTLV                                 270

Qy  359  QRPPRPASHPQASFNLASPEKVSNTTVV  386
```

```
RESULT 3
ID   Q9DF61        PRELIMINARY;      PRT;   344 AA.
AC   Q9DF61;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   OBCAM alpha 1 isoform.
GN   OBCAM.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=20499204; PubMed=11042360;
RA   Lodge A.P., Howard M.R., McNamee C.J., Moss D.J.;
RT   "Co-localisation, heterophilic interactions and regulated expression
RT   of IgLON family proteins in the chick nervous system.";
RL   Brain Res. Mol. Brain Res. 82:84-94(2000).
DR   EMBL; AF292934; AAG01877.1; -.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 2.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
KW   Immunoglobulin domain.
SQ   SEQUENCE   344 AA;  37531 MW;  37F6051CBF0E7B4 CRC64;
```

RESULT 4
Q9D912

Query Match
Best Local Similarity 26.8%; Pred. No. 0.00014;
Matches 72; Conservative 35; Mismatches 110; Indels 59; Gaps 15;

Score 142.5; DB 13; Length 344;

```
Qy    3 RHLLTVPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCIVSQGWKLIMWALSDMVVL---  58
Db   20 RLLFLVPAGVPVRSGDATFPKRAMDNVTVRQGESATLRCTVYDDRVRRVAW--LNRSTILYAG  78

Qy   59 ---SVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSR  108
Db   79 NDKWSIDNRVVLLSN--TKTQY---------------SIKIHNVDVYDEGPYTCSVQTDNHPKTSR  127

Qy  109 LHGSAYLTVQVWGELFIPSVNLVVAENEPCEVTCL-----PSHWTRLPDISWELGLLVSHS  164
Db  128 VH----LIVQVPQIVNISSDITVNEGSSVTLMCLAFGRPE---PTVTWR--HLSGK  175

Qy  165 SYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-----IRCPQD  219
Db  176 GQGFVSEDEYLE---ITGITRREQSGEYECSAV-NDVAPVDVRKVTVNYPPYISNAKN  230

Qy  220 TGGGINIPGVL----SSLPSLGFSLPTWGKVGLGLA  251
Db  231 TGASVGQKGILQCEASAVPVAEFQ---WFKEDTRLA  263
```

ID   Q9D912
AC   Q9D912;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   4931420D14Rik protein.
GN   4931420D14Rik.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK006992; BAB24782.1; -.
DR   MGD; MGI:1919992; 4931420D14Rik.
SQ   SEQUENCE 168 AA; 18931 MW; 7A2BD279612A5E94 CRC64;

PRELIMINARY; PRT; 168 AA.

RESULT 5
Q9DGI5

Query Match
Best Local Similarity 26.8%; Pred. No. 0.00036;
Matches 70; Conservative 38; Mismatches 104; Indels 57; Gaps 14;

Score 137.5; DB 13; Length 315;

```
Qy    3 RHLLTVPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCIVSQ----GW--KLIMWALS  53
Db   20 RLLFLVPAGVPVRSGDATFPKRAMDNVTVRQGESATLRCIVYDDRVRVAWLNRSSTLYAGN  79

Qy   54 DMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ----NS  107
Db   80 DKWCLDPRVLLANTKTQYSIQ-------------------IHDVDVYDEGPYTCSVQTDNHPKTS  126

Qy  108 RLHGSAYLTVQVWGELFIPSVNLVVAENEPCEVTCLPSHWTRLPD--ISWELGLLVSHS  165
Db  127 RVH----LIVQVSPKITEISSDISINEGGNVSLICIA--TGRPDPTITWR--HISPKA  176

Qy  166 YYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV----IRCPQDT  220
Db  177 VGFISEDEYLE---ITGITREQSGEYECSAS-NDVAAPVVQRKVTVNYPPYISDAKST  231

Qy  221 GGGINIPGVL----SSLPSLGFSLPTWGK  245
Db  232 GVPVGQKGILMCEASAVPSADFQ---WTK  257

Qy  315 NENSGYNSDEQKTETASLPPKSCESSSDPEQRNSSCGPPHQRADQRPRPPASHPQASFNL  374
Db  112 LTTRSQLELIEPE--PIMALEPSEITVAFFSHKNANVSDP-----EEVPPCLDSDPFPNGDL  165

Qy  375 AS  376
Db  166 AS  167
```

ID   Q9DGI5
AC   Q9DGI5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CEPU-Se alpha 1 isoform.
GN   CEPU-SE.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Lodge A.P., McNamee C.J., Howard M.R., Reed J.E., Moss D.J.;
RT   "Characterisation of CEPU-Se, a secreted isoform of the IgLON family
RT   protein CEPU-1.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF292936; AAG01879.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain.
SQ   SEQUENCE 315 AA; 34606 MW; 68C5D27F0DDC6FB2 CRC64;

PRELIMINARY; PRT; 315 AA.

RESULT 6
Q99PJO

Query Match
Best Local Similarity 32.0%; Pred. No. 7.7e-05;
Matches 39; Conservative 20; Mismatches 49; Indels 14; Gaps 6;

Score 141; DB 11; Length 168;

```
Qy  260 CTLTLTRCCCCRRRCGCGNC-CCRCFCCCRRKRGFR---IQFQKKSEKETNKEIETESG  314
Db   55 CSLSRSCCCCRCRCC-CHCRCRC--CCSRSRRFRSRRTLKRFQITEKGEQSLQRRIRRQ  111
```

ID   Q99PJO
AC   Q99PJO;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)

DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Neurotrimin.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR; TISSUE=BRAIN;
RA   Kim T.H., Choi S.C., Kim J., Jeon J.W., Kim K.D., Lee S.H.;
RT   "Cloning and expression of mouse neurotrimin gene in the developing
RT   nervous system.";
RL   Submitted (JUN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF282980; AAK00276.1; -.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00409; IG; 3.
DR   SMART; SM00410; IG_like; 2.
KW   Immunoglobulin domain.
SQ   SEQUENCE   344 AA; 37924 MW; 3ECC6D5EE6C5C17D CRC64;

Query Match   6.6%; Score 137.5; DB 11; Length 344;
Best Local Similarity   26.0%; Pred. No. 0.0004;
Matches   68; Conservative   36; Mismatches 115; Indels   43; Gaps   13;

Qy   3 RHLLIVPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVR 61
Db   20 RLLFLVPIGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDNRVTRVAWLNRSTI----- 74

Qy   62 PMEPIITNDRF-TSQRYDQGNFTSE--MIIHNVEPSDSGNIRCSLQ------NSRLHGS 112
Db   75 ---PYAGNDKWCLLDPRVLLGNTCTQYSIEIQNVDVDEGPYTCSVQIDNHPKTSRVH-- 129

Qy 113 AYLTVQVWGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEP 172
Db 130 --LIVQVSPKIVEISSDISINEGNNISLTCIATGRPE-PTVTWR--HISPKAVGFVSED 183

Qy 173 SDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP------QDTGGGINIP 227
Db 184 EYLQ----IQGITREQSGEYECSAS-NDVAAPVVPRVKVTVNPPYFSEAKGTGVPVGQK 238

Qy 228 GVL-----SSLPSLGFSLPTWGK 245
Db 239 GTLQCERASAVPSAEFQ---WFK 257

RESULT 7
O93242
ID   O93242   PRELIMINARY;   PRT;   344 AA.
AC   O93242;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CEPU-1.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kimura Y., Shirabe K., Fukushima M., Takeshita M., Tanaka H.;
RT   "CEPU-1; an Immunoglobulin Superfamily Molecule, Has Cell Adhesion
RT   Activity and Shows Dynamic Expression Patterns in Chick Embryonic
RT   Spinal Cord.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB011810; BAA31514.1; -.
DR   InterPro; IPR003598; Ig_C2.

RESULT 8
Q9D4K2
ID   Q9D4K2   PRELIMINARY;   PRT;   173 AA.
AC   Q9D4K2;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   4931420D14Rik protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=TESTIS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK016467; BAB30253.1; -.
DR   MGD; MGI:1913992; 4931420D14Rik.
SQ   SEQUENCE   173 AA; 19581 MW; 147B6F155AC29FDF CRC64;

Query Match   6.6%; Score 137.5; DB 13; Length 344;
Best Local Similarity   26.0%; Pred. No. 0.0004;
Matches   70; Conservative   38; Mismatches 104; Indels   57; Gaps   14;

Qy   3 RHLLIVPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCTVSQ-----GW--KLIMWALS 53
Db   20 RLLFLVPAGVPVRSGDATFPKAMDNVTVRQGESATLRCSVDNRVTRVAWLNRSSILYAGN 79

Qy   54 DMVVLSVRPMEPIITNDRFTSQRYDQGNFTSEMIIHNVEPSDSGNIRCSLQ------NS 107
Db   80 DKWCLDPRVVLLANTKTQYSIQ----------IHDVDVYDEGPYTCSVQTDNHPKTS 126

Qy 108 RLHGSAAYLTVQVWGELFIPSVNLVVAENEPCEVTCLPSHWTRLPD--ISWELGLLVSHSS 165
Db 127 RVH-------LIVQVSPKIEISSDISINEGGNVSLTCIA---TGRPDPITTWR--HISPKA 176

Qy 166 YYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTIV------IRCPDT 220
Db 177 VGFISEDEYLE----ITGTTREQSGEYECSAS-NDVAAPVVQRVKVTVNYPPYISDAKST 231

Qy 221 GGGINIPGVL-----SSLPSLGFSLPTWGK 245
Db 232 GVPVGQKGILMCEASAVPSAEFQ---WYK 257

RESULT 9

Q92626

```
Query Match             6.6%;   Score 136.5;   DB 11;   Length 173;
Best Local Similarity  31.5%;   Pred. No. 0.0002;
Matches  40;  Conservative 16;  Mismatches 52;  Indels 19;  Gaps  6;

Qy  260 CTLTIRCCCRRRCCGCNC--CCRCCFCCRRKRGFRIQFOKKSEKEKTNKETESGNENS 318
        |:|:|  ||||| ||||  | |||| ||  ||:||| ||| |||||| |||||||
Db   55 CSLSRSCCCCRCC-CYCRCRCRC--CCSRSRRFRSRTTLRVRDPKFFQ--ITEKGEQSL 109

Qy  319 GYNSDEQKTTE------TASLPPKSCCESSDPEQRNSSCGPPHQRADQRPRPASHPQ 369
        |||| |||||                                          |||||
Db  110 QRRIRQLTRSQLELIEPEPTMALEPSEITVAFFSHKNANVSDP-----EEVPCLDSDPF 165

Qy  370 ASFNLAS 376
        : :|||
Db  166 PNGDLAS 172
```

```
ID   Q92626    PRELIMINARY;   PRT;   1496 AA.
AC   Q92626;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   MYELOBLAST KIAA0230 (Fragment).
GN   KIAA0230.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BONE MARROW;
RX   MEDLINE=97191544; PubMed=9039502;
RA   Nagase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O.,
RA   Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT   "Prediction of the coding sequences of unidentified human genes. VI.
RT   The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by
RT   analysis of cDNA clones from cell line KG-1 and brain.";
RL   DNA Res. 3:321-329(1996).
DR   EMBL; D86983; BAA13219.1; -.
KW   Immunoglobulin domain.
FT   NON_TER       1      1
SQ   SEQUENCE   1496 AA;  167209 MW;  E9B9A7069EF1ABFF CRC64;
```

```
DR   Pfam; PF00093; vwc; 1.
DR   PRINTS; PR00457; ANPEROXIDASE.
DR   SMART; SM00408; IGc2; 4.
DR   SMART; SM00082; LRRCT; 1.
DR   SMART; SM00013; LRRNT; 1.
DR   SMART; SM00369; LRR_TYP; 4.
DR   SMART; SM00214; VWC; 1.
DR   PROSITE; PS01208; VWFC; UNKNOWN_1.
```

RESULT 10

Q9P121

```
ID   Q9P121    PRELIMINARY;   PRT;   344 AA.
AC   Q9P121;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Neurotrimin.
HNT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Li G., Jin J., Tan X., Hu S., Yuan J., Qiang B.;
RT   "Cloning and identification of human neurotrimin full length cDNA.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF126426; AAF37591.1; -.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   Pfam; PF00047; Ig; 3.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00410; IG_like; 1.
KW   Immunoglobulin domain.
SQ   SEQUENCE   344 AA;  37971 MW;  DA4D12C295ABBE3A CRC64;
```

```
Query Match            6.5%;   Score 135;   DB 4;   Length 1496;
Best Local Similarity 25.0%;   Pred. No. 0.0046;
Matches  73;  Conservative 36;  Mismatches 127;  Indels 56;  Gaps 15;

Qy   13 GSGSGNEVIEGPQNATVLKGSQARFNCTVS-QGNKLIMWALSDMVVLSVRPMEPIITNDR 71
        |||  |||  |||| ||| ||  | ||  :   ||  | |||  |   ||:  :||||
Db  354 GSPARPTEVIQPQNTEVLVGESVTLECSATGHPPPRISWTRGDRTPLPVDPRVNITPS-- 411

Qy   72 FTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNS--RLHGSAYLTVQVMGELFIPSVN 129
        |     | ||     |  |  |  |||  | |   |   | ||| ||||  |||
Db  412 --------GG----LYIONVVQGDSGEYACSAINNIDSVHATAFIIVQALPQFTVTPQD 458

Qy  130 LVVAENEPCEVTCLPFSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQS 188
        |||   :  :  | |     |    :|   |  ||   | ||||||  |  || :  |
Db  459 RVVIEGQTVDFQC-EAKGNPPPVIAWTKGGSQLSVDRRHLVLSSGTLR--ISGVALHDQ- 514

Qy  189 NGTLTCVATWKSLKARKSATVNLTV---------TRCPQDT----GGGINIPGVLSSLPSL 236
        : |  |  |   :      | :|:         ||| ||       ||||| |   | |
Db  515 -GQYECQAV--NIIGSQKVVAHLTVQPRVTPVFASIPSDTTVEVGANVQLP-----CSSQ 566

Qy  237 GFSLP---TWGKVGLGL--AGTMLLTPTCTLTI--------RCCCCRRCCG 275
        |       ||  |       | | | |  | :         || |
Db  567 GEPEPAITWNKDGVQVTESGKFHISPEGFLTINDVGPADAGRYECVAARNTIG 618
```

```
Query Match            6.4%;   Score 133.5;   DB 4;   Length 344;
Best Local Similarity 24.5%;   Pred. No. 0.00092;
Matches  65;  Conservative 40;  Mismatches 111;  Indels 49;  Gaps 13;

Qy    3 RHLLTVPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCTVSGQWKLIMWALSDMVVLS-- 59
          ||:| |  ::|| ||||| ||||||||||||||||||||||||||||||||||||
Db   20 RLLFLVPTGVPVRSGDATFPKAMDNVIVRQGESAILRCTIDNRVTRVAWLNRSTILYAGN 79
```

```
Qy   60 -----VRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGGNIRCSLQ------NSRL 109
            :::              |:|                          |::        |:|
Db   80 DKWCLDPRVVLLSN---TQTQY------SIEIQNVDVYDEGPYTCSVQTDNHPKTISRV 128

Qy  110 HGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFV 169
            :|                                  |||               :|
Db  129 H----LIVQVSPKIVEISSDISINEGNNISLTCIATGRPE-PTVTWR---HISPKAVGFV 180

Qy  170 PEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP-----QDIGGGI 224
            |:|                                                    ||:
Db  181 SEDEYLE----IQGITREQSGDYECSAS-NDVAAPVVRRVKVTVNPPYISEAKGTGVPV 235

Qy  225 NIPGVL-----SSLPSLGFSLPTWGK 245
            |::|        |::|        |:|
Db  236 GQKGTLQCEASAVPSAEFQ---WYK 257
```

RESULT 11

```
ID   Q91562      PRELIMINARY;      PRT;   1427 AA.
AC   Q91562;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Tumor suppressor.
GN   XDCCA.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9513183; PubMed=7813784;
RA   Pierceall W.E., Reale M.A., Candia A.F., Wright C.V., Cho K.R.,
RA   Fearon E.R.;
RT   "Expression of a homologue of the deleted in colorectal cancer (DCC)
RT   gene in the nervous system of developing Xenopus embryos.";
RL   Dev. Biol. 166:654-665(1994).
DR   EMBL; U10996; AAA70168.1; -.
DR   HSSP; P40189; 1BQU.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   Pfam; PF00047; ig; 4.
DR   Pfam; PF00041; fn3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
DR   SMART; SM00060; FN3; 6.
DR   PRINTS; PR00014; FNTPEIII.
DR   PRINTS; PR00011; 1BQU.
KW   Immunoglobulin domain; Repeat.
SQ   SEQUENCE 1427 AA; 156533 MW; 61FEA12C8A674972 CRC64;

Query Match     6.4%;   Score 133;  DB 13;  Length 1427;
Best Local Similarity   25.7%;  Pred. No. 0.0065;
Matches  59;  Conservative  33;  Mismatches  90;  Indels  48;  Gaps  13;

Qy   21 IEGPQNATVLKGSQARFNCTVSQGW--KLIMWALSDMVVLSVRMEPIITNDRFTSQRVD 78
            :::|||||    ||||||||||||   |||||||||||||||||||||||||||||:
Db  243 LQRPSNVVAIEGQDAVLECAVS-GYPTPIVVMQGD---------EPVPIRTR---KYS 288

Qy   79 QGGNFTSEMIIHNVEPSDSGGNIRC--SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAENE 136
            |||||||||||||||||||||||    |||||||||||||||||||||||||||||||
Db  289 VLGG--SNLLISNVTDDDAGAYTCVATYKNENTSFSADLTVMVPPQFLNHPANLYAYESM 346

Qy  137 PCEVTCL---PSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGT 191
            |::|:||     ||    |      |:|||:||||:||||||||||||||||||||||
Db  347 DIEFECAVSGKPS------PTVKWTKNGEVVIPSDYFQIVDGSNIR----ILGLVKSDEGY 397

Qy  192 LTCVA-------TWKSLKARKSATVNLTVI-RCPQDTGGGINIPGVLSS 232
            |:|          ||:|||||:||||| :|  :|   |||||:||::
Db  398 YQCIAENEAGNIQTYAQLIIPDPAVPSSSILPSAPRDV-----VPVLVSS 442
```

RESULT 12

```
ID   Q9DBP0      PRELIMINARY;      PRT;   697 AA.
AC   Q9DBP0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Solute carrier family 34 (sodium phosphate), member 2.
GN   SLC34A2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=LUNG;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK004832; BAB23600.1; -.
DR   MGD; MGI:1342284; Slc34a2.
DR   InterPro; IPR003841; Na_Pi_cotrans.
DR   Pfam; PF02690; Na_Pi_cotrans; 1.
DR   TIGRFAMs; TIGR01013; 2a58; 1.
SQ   SEQUENCE 697 AA; 76244 MW; 2A7B9384857EF16F CRC64;

Query Match     6.3%;   Score 131;  DB 11;  Length 697;
Best Local Similarity   20.3%;  Pred. No. 0.0039;
Matches  88;  Conservative  53;  Mismatches  124;  Indels  168;  Gaps  22;

Qy    5 LLTVPEAVGSGSGNEVIEGPQNATVLK------------GSQARFNCTVS 42
           :|:|||                |:||:             |||||||:
Db  380 LIMIVKLLGS-----VLRG-QVAPVIKKTLNTDPFPPAWLTGYLAILVGAGWTFIVQSS 433

Qy   43 QGWKLIMWALSDMVVLSVRMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGGNIRC 102
           | | ||||||||||||||||||||||||||      ||||||||::            :
Db  434 SVFTSAMTPLIGIGVISTERAYPLTIGSNI------GTTTTAILAALASPGNT--LRS 483

Qy  103 SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLP-------DIS 154
           |||   |||||||||||||||||||||     |          |:|     ||
Db  484 SLQIALCH----------------------FFFNISGI-----LLWYPIPFTWTRLPIRLAKGLGNIS 522

Qy  155 -----WELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVN 210
              || ||||||||||:|||||||||||||||||||| |   |::| |||:|||||:  |
Db  523 AKFRW--PAVFLIFFFVTP-----LTVFGLS------LAGWPVLVGVGVPIIL 564

Qy  211 LTVI---------RCPQDTGGGINIPGVL-----SSLPSLGFSLPTWGKVGLGLAGTMLLTP 258
           |:|           :|   ||||||||  |      :::|   | ||| |  |   |||
Db  565 LILLVICLLRMLQFRCPR----ILPLKLRDVNFPLLMMHSLKPWDNV-------ISLAT 611

Qy  259 TCTLTIRCCCCRRRCCG------CNC-CCRCFCCRRKRGFRIQFQKRSEKEKTNKETETE 312
           || || |  |  | |||       |:   | ||   |:||:| |::| :| |::
```

```
Db  612  TC-FQRRCCCCRVCCRVCCMVCGKCCCRGSKCCCRD-----...-----QGEEEEKEQ-----  656
              :||: :||:|| | |                        :||:||||
Qy  313  SGNENSGYNSDEQKTTETASLPPKSCCESSSDPEQRNSSCGPPHQRADQRPPRPASHPQASF  372

Qy  657  -----DIPVKASGAFDNAAMSKEC-----QDEGKGQV  683
              :|: | |||||                  ||||||
Db  684  EVLSMKALSNTTV  696

Qy  373  NLASPEKVSNTTV  385
```

RESULT 13
Q92290

```
ID   Q92290      PRELIMINARY;      PRT;   697 AA.
AC   Q92290;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Type IIb Na/phosphate-cotransporter.
GN   SLC34A2 OR NPT2B.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NMRI; TISSUE=SMALL INTESTINE;
RX   MEDLINE=99045724; PubMed=9827740;
RA   Hilfiker H., Hattenhauer O., Traebert M., Forster I., Murer H.,
RA   Biber J.;
RT   "Characterization of a murine type II sodium-phosphate cotransporter
RT   expressed in mammalian small intestine.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:14564-14569(1998).
DR   EMBL; AF081499; AAC80007.1; -.
DR   MGD; MGI:1342284; Slc34a2.
DR   InterPro; IPR003841; Na_Pi_cotrans.
DR   Pfam; PF02690; Na_Pi_cotrans; 1.
DR   TIGRFAMs; TIGR01013; 2a58; 1.
SQ   SEQUENCE   697 AA;  76286 MW;  839E5CCBF0F565265 CRC64;
```

```
Query Match             6.3%;  Score 131;  DB 11;  Length 697;
Best Local Similarity  20.3%;  Pred. No. 0.0039;
Matches   88;  Conservative   53;  Mismatches 124;  Indels 168;  Gaps 22;

Qy    5  LLTVPEAVGSGSGNEVIEGPQNATVLK-----...-----GSQARFNCTVS   42
Db  380  LIMIVKLLGS-----VLRG-QVATVIKKTLNTDPFPFAWLTGYLAILVGAGMTFIVQSS  433

Qy   43  QGWKLIMWALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC  102
Db  434  SVFTSAMTPLIGIGVISIERAYPLTLGSNI-----...-----GTTTTAILAALASPGNT--LRS  483

Qy  103  SLQNSRLHGSAVITVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLP-------DIS  154
Db  484  SLQIALCH-----...-----FFFNISGI  522

Qy  155  -----WELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVN  210
Db  523  AKYRW--FAVFYLIFFFVFD-----LTVFGLS-----LAGWPVLVGVGVPIIL  564

Qy  211  LTVI-----RCPQDTGGGINIPGVL-----SSLPSLGFSLPTWGKVGLGLAGTMLLTP  258
Db  565  LILLVICLRMLQFRCPR-----ILPLKLRDWNFLPLWMHSLKPWDNV-----ISLAT  611

Qy  259  TCTLTIRCCCCRRRCCG-----CNC-CCRCCFCCRRKRGFRIQFQKKSEKEKTNKETETE  312
Db  612  TC-FQRRCCCCRVCCRVCCMVCGKCCCRGSKCCCRD-----...-----QGEEEEKEQ-----  656

Qy  313  SGNENSGYNSDEQKTTETASLPPKSCCESSSDPEQRNSSCGPPHQRADQRPPRPASHPQASF  372
Qy  657  -----DIPVKASGAFDNAAMSKEC-----QDEGKGQV  683
Db  684  EVLSMKALSNTTV  696
Qy  373  NLASPEKVSNTTV  385
```

```
Db  684  EVLSMKALSNTTV  696
```

RESULT 14
Q22048

```
ID   Q22048      PRELIMINARY;      PRT;   164 AA.
AC   Q22048;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   T01B7.8 protein.
GN   T01B7.8.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sims M.A.;
RL   Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z66499; CAA91301.1; -.
DR   InterPro; IPR001450; 4Fe4S_ferredoxin.
DR   InterPro; IPR001271; Defensin_mammal.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001007; VWF_C.
DR   PROSITE; PS00198; 4FE4S_FERREDOXIN; UNKNOWN_1.
DR   PROSITE; PS00199; DEFENSIN; UNKNOWN_1.
DR   PROSITE; PS00269; EGF_1; UNKNOWN_1.
DR   PROSITE; PS00022; VWFC; UNKNOWN_1.
DR   PROSITE; PS01208; EGF_1; UNKNOWN_1.
SQ   SEQUENCE   164 AA;  16499 MW;  C002D48D36C9FCED CRC64;
```

```
Query Match             6.2%;  Score 129.5;  DB 5;  Length 164;
Best Local Similarity  32.1%;  Pred. No. 0.00081;
Matches   43;  Conservative   11;  Mismatches  53;  Indels 27;  Gaps 6;

Qy  157  LGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRQ  216
Db    6  LAILLAIGTFIAV--SQVQSAV-----LPVSSTELATVGTDVSTASTAIDTLGNSSSRV   57

Qy  217  PQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCLTIRCCCCRRRCCGC  276
Db   58  KRQGGCGCGGCGC-----GCCGCGGGGGG--CGCCCCRPRCCCCCRRCCTC  101

Qy  277  -NCCC-RCCFCCR  287
Db  102  CRPCCCTRCCTCCR  115
```

RESULT 15
001761

```
ID   001761      PRELIMINARY;      PRT;   6632 AA.
AC   001761;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   C. elegans UNC-89 protein (corresponding sequence C09D1.1).
GN   UNC-89.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
```

```
RA   Waterston R.;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Du Z., Le T.T., Wilson R.;
RT   "The sequence of C. elegans cosmid C09D1.";
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RL   Submitted (APR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF003131; AAB54132.2; -.
SQ   SEQUENCE   6632 AA;  731665 MW;  262D3EDD62960E89 CRC64;

Query Match
Best Local Similarity   23.0%;   Score 129;  DB 5;  Length 6632;
Matches  59;  Conservative  41;  Mismatches  97;  Indels  60;  Gaps  10;

Qy      6 LTVPEAVGSG--------SGNE------------------VIEGPQNATVLKGSQARFNCI 40
            : :|  |   | ||                           |::||:| |:      :
Db   2038 LIIPNAQDSGKITVEASNEVGSSESSAQLTVNPSTTPIVVDGPKSVTIKETETAEFKAT 2097

Qy     41 VSQGW--KLIMWALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSG 98
            :||    :| | |:||||  : | |  | | : || ::|||:|| :||||::    ||
Db   2098 IS-GFPAPTVKWTINEKIVEESRITITIKTEDVYT-----------LKISNAKIEQTG 2143

Qy     99 NIRCSLQNSRLHGSAYLTVQVMGELFIPS------VNLVVAENEPCEVTCL---PSHWTRL 150
            :: ||| | |:  : | |   : |:|           : | :| | :    :  |: | :|
Db   2144 TVKVTAQNSAGQDSKQADLKVEPNVKAPKFKSQLTDKVADEGEPLRWNLELDGPSPGT-- 2201

Qy    151 PDISWEL-GLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATV 209
            | :|| |  | |: || : | |  : || |: |::|  ||| | |     |||:|||| |
Db   2202 -EVSWLLNGQPLTKSDTVQVVDHGDGTYHVTIAEAKPEMSGTLTAKAKNAAGECETSAKV 2260

Qy    210 NLT--------VIRCPQD 219
            : |        || |:
Db   2261 TVNGGNKKPEFVQAPQN 2277

Search completed: April 28, 2003, 21:07:42
Job time : 41.1283 secs
```

OM protein - protein search, using sw model

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 16:24:30 ; Search time 10.3692 Seconds
                                           (without alignments)
                                           1543.990 Million cell updates/sec

Title: US-09-729-264-6
Perfect score: 2077
Sequence: 1 MERHLLTVPEAVGSGSGNEV............HPQASFNLASPEKVSNTTVV 386

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB	ID	Description
1	151	7.3	404	1	RAGE_HUMAN	Q15109 homo sapien
2	139.5	6.7	344	1	NTRI_RAT	Q62718 rattus norv
3	133	6.4	337	1	G55A_CHICK	Q98892 gallus gall
4	131	6.3	1447	1	DCC_MOUSE	P70221 mus musculu
5	130	6.3	345	1	OPCM_HUMAN	Q14882 homo sapien
6	129	6.2	345	1	OPCM_BOVIN	P11834 bos taurus
7	129	6.2	353	1	CEPU_CHICK	Q90773 gallus gall
8	128	6.2	345	1	OPCM_RAT	P32736 rattus norv
9	128	6.2	1447	1	DCC_HUMAN	P43146 homo sapien
10	128	6.2	1914	1	KMLS_HUMAN	Q15746 homo sapien
11	125	6.0	1070	1	PTK7_HUMAN	Q13308 homo sapien
12	122.5	5.9	4393	1	PGBM_HUMAN	P98160 homo sapien
13	118.5	5.7	620	1	SMP_COTJA	Q92154 coturnix co
14	117.5	5.7	1443	1	NEOL_CHICK	Q90610 gallus gall
15	114.5	5.5	1461	1	NEOL_HUMAN	Q92859 homo sapien
16	113.5	5.5	416	1	RAGE_BOVIN	Q28873 bos taurus
17	113.5	5.5	1377	1	NEOL_RAT	P97603 rattus norv
18	111	5.3	365	1	CXAR_HUMAN	P78310 homo sapien
19	110.5	5.3	524	1	BUTY_MOUSE	Q62556 mus musculu
20	110.5	5.3	764	1	ICCR_DROME	Q08180 drosophila
21	109.5	5.3	249	1	CSP_DROME	Q03751 drosophila
22	109.5	5.3	890	1	TYO3_HUMAN	Q06418 homo sapien
23	106	5.1	1092	1	NCA2_XENLA	P36335 xenopus lae
24	105.5	5.1	862	1	CD22_MOUSE	P35329 mus musculu
25	105.5	5.1	3707	1	PGBM_MOUSE	Q05793 mus musculu
26	105	5.1	319	1	A33_HUMAN	Q99795 homo sapien
27	104.5	5.0	333	1	AMAL_DROME	P53364 drosophila
28	104.5	5.0	879	1	FPRP_RAT	Q62786 rattus norv
29	104.5	5.0	338	1	LAMP_CHICK	Q99919 gallus gall
30	102.5	4.9	344	1	CD2_RAT	P08921 rattus norv
31	102.5	4.9	348	1	KILO_RAT	Q92J8 rattus norv
32	102.5	4.9	365	1	CXAR_MOUSE	P97792 mus musculu
33	100.5	4.8	830	1	SREC_HUMAN	Q14162 homo sapien
34	100.5	4.8	1906	1	KMLS_CHICK	P11799 gallus gall
35	100.5	4.8	2481	1	UN52_CAEEL	Q06561 caenorhabdi
36	100	4.8	873	1	FAS2_DROME	P34082 drosophila
37	100	4.8	1010	1	CONT_CHICK	P14781 gallus gall
38	100	4.8	1241	1	NPHN_HUMAN	O60500 homo sapien
39	99.5	4.8	880	1	TYO3_RAT	O08500 rattus norv
40	99.5	4.8	1260	1	CAML_MOUSE	P55146 mus musculu
41	99	4.8	338	1	LAMP_HUMAN	P11627 homo sapien
42	99	4.8	338	1	LAMP_RAT	Q13449 rattus norv
43	99	4.8	359	1	LACH_DROME	Q62813 drosophila
44	99	4.8	2499	1	MPRI_BOVIN	Q24372 bos taurus
45	98.5	4.7	858	1	NCA1_RAT	P08169 rattus norv

ALIGNMENTS

RESULT 1
RAGE_HUMAN

ID   RAGE_HUMAN   STANDARD;   PRT;   404 AA.
AC   Q15109; Q15279; Q9Y3R3; Q9H2X7;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Advanced glycosylation end product-specific receptor precursor
DE   (Receptor for advanced glycosylation end products).
GN   AGER OR RAGE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=92340547; PubMed=1378843;
RA   Neeper M., Schmidt A.M., Brett J., Yan S.D., Wang F., Pan Y.C.,
RA   Elliston K., Stern D., Shaw A.;
RT   "Cloning and expression of a cell surface receptor for advanced
RT   glycosylation end products of proteins.";
RL   J. Biol. Chem. 267:14998-15004(1992).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Lung;
RX   MEDLINE=95137587; PubMed=7835890;
RA   Sugaya K., Fukagawa T., Matsumoto K., Mita K., Takahashi E., Ando A.,
RA   Inoko H., Ikemura T.;
RT   "Three genes in the human MHC class III region near the junction with
RT   the class II: gene for receptor of advanced glycosylation end
RT   products, PBX2 homeobox gene and a notch homolog, human counterpart
RT   of mouse mammary tumor gene int-3.";
RL   Genomics 23:408-419(1994).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Rowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA   Banta A., Spies T., Hood L.;
RL   Submitted (FEB-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT ARG-100.
RA   Abedin M.J., Yonekura H., Migita H., Karasawa J., Yamamoto Y.,
RA   Yamamoto H.;
RT   "Molecular heterogeneity of the receptor for advanced glycation
RT   endproducts.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RA   Malherbe P., Richards J., Gaillard H., Thompson A., Diener C.,
RA   Schuler A., Huber G.;
RT   "cDNA cloning of a novel secreted isoform of the human Receptor for
RT   Advanced Glycation End products (RAGE) and characterization of cells
RT   co-expressing cell-surface scavenger receptors and Swedish mutant
RT   amyloid precursor protein.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM 1).

RC   TISSUE=Lung;
RA   Strausberg R.;
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [7]
RP   SEQUENCE OF 1-12 FROM N.A.
RA   Hudson B.I., Futers T.S.;
RL   "Novel polymorphisms in the receptor for advanced glycation
RL   end-products (RAGE) gene.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MEDIATES INTERACTIONS OF ADVANCED GLYCOSYLATION END
CC   PRODUCTS (AGE). THESE ARE NONENZYMATICALLY GLYCOSYLATED PROTEINS
CC   WHICH ACCUMULATE IN VASCULAR TISSUE IN AGING AND AT AN ACCELERATED
CC   RATE IN DIABETES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein (isoform 1).
CC   Secreted (isoform 2).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2/RAGESEC;
CC   ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ENDOTHELIAL CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; M91211; AAA03574.1; -.
DR   EMBL; D28769; BAA05958.1; -.
DR   EMBL; U89336; AAB47491.1; -.
DR   EMBL; AB036432; BAA89369.1; -.
DR   EMBL; AJ133822; CAB43108.1; -.
DR   EMBL; BC020669; AAH20669.1; -.
DR   EMBL; AF202289; AAG35728.1; -.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   MIM; 600214; -.
DR   Genew; HGNC:320; AGER.
KW   Immunoglobulin domain; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Alternative splicing; Polymorphism.
FT   SIGNAL   1   22   POTENTIAL.
FT   CHAIN   23   404
FT   ADVANCED GLYCOSYLATION END PRODUCT-
FT   SPECIFIC RECEPTOR.
FT   DOMAIN   23   342   EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   343   363   POTENTIAL.
FT   DOMAIN   364   404   CYTOPLASMIC (POTENTIAL).
FT   DOMAIN   31   106   IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN   137   215   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN   252   308   IG-LIKE C2-TYPE DOMAIN 2.
FT   DISULFID   38   99   POTENTIAL.
FT   DISULFID   144   208   POTENTIAL.
FT   DISULFID   259   301   POTENTIAL.
FT   CARBOHYD   25   25   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   81   81   N-LINKED (GLCNAC...) (POTENTIAL).
FT   DOMAIN   380   384   POLY-GLU.
FT   VARSPLIC   54   67   MISSING (IN ISOFORM 2).
FT   VARSPLIC   275   404
FT   GVPLPLPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRA
FT   VSISIIEPGEEGPTAGSVGGSGGLGTLALALGILGGLGTAAL
FT   LIGVILWQRQRQRGEERKAPENQEEEEERAELNQSEEPEAG
FT   ESSTGGP -> VSDLERGAGRTRPGGANCRLCGRIRAGNSS
FT   PGPGDPGPGDSRPAHWGHLVAKAAATPRRGEEGPRKPGGRG
FT   GACRTESVGGT (IN ISOFORM 2).
FT   VARIANT   100   100   Q -> R.
FT   /FTId=VAR_011338.
FT   CONFLICT   1   1   M -> G (IN REF. 1).

SQ   SEQUENCE   404 AA; 42802 MW; 0D58AC4436C30CCE7 CRC64;
Qy   Query Match 90; Conservative 38; Mismatches 112; Indels 146; Gaps 19;
     Best Local Similarity 23.3%; Pred. No. 7e-05;
     Matches 90;   Score 151;   DB 1; Length 404;

Qy   9 PEAVGSGSGCNEVIEGPQNATVLKGSQARFNCTVSQG--WKLIMWALSDMVVLSVRPMEP 65
             |:|:   ::| ||:| |:|||||
Db   124 PEIVDSAS--ELRAGVPN----------KVGTCVSEGSYPAGTLSWHLDG------KP 163

Qy   66 IITNDRFTS-----QRYDGGNFT--SEMIIHNVEPSDSGNIR---CSLQNSRLHGSAY 114
             |:|:: |::.::||| |:|:|:| ||: :||: |:|::|||:||:|
Db   164 LVPNEKGVSVKEQTRRHPETGLFTLQSELM--VTPARGGDRPTFSCSFSPGLPRHRAL 220

Qy   115 LTVQVMGELFIP----SVNLVVAENEP----------------CEVTCLPSHWTRLPDISWE 156
             |:|| | | ||||||| |:| :|| :|| :|
Db   221 RTAPIQPRVWEPVLEEVQLVV---EPEGGAVAPGGTVTLTCEVPAQPS----PQIHWM 272

Qy   157 LGLLVSHSSYYVFEPSDLQSAVSILALTPQSNGTLTCVATWRSLKARKSATVNLTVIRC 216
             || |::::|| :|:| || |:| :|| ||||:|||:|:| |::|:|:|| :|
Db   273 KD--------GVPLPLPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIE- 322

Qy   217 PQDTGGGINIPGVLSSLPSLGFSLPTWGRKVGLGLAGTMLLT-------PTCLTIRCCCC 269
             |::||||| :| | |:| || :|:| |||:|||| ||||| |||||||||||
Db   323 PGEEG----------PTAGSVGGSGLGTLALALGILGGLGTAALLIGVILVIW 363

Qy   270 RRCCGCGNCCCRCCCRKCRRKRGFRIQFQKKSEKEKT--NKETFTESGNENSGYNSDEQKT 327
             |:|:|:| :|:| ::|::| ||:||: |:|:
Db   364 QRR------------QRRGEERKAPENQEEEEERAELN------ 389

Qy   328 TETASLPPKSCESSDPEQRNSSCGPP 353
             :|:||| |:||
Db   390 -----------------QSEEPEAGESSTGGP 404

RESULT 2
NTRI_RAT
ID   NTRI_RAT   STANDARD;   PRT;   344 AA.
AC   Q62718;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Neurotrimin precursor (GP65).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 217-229.
RX   MEDLINE=95198094; PubMed=7891157;
RA   Struyk A.F., Canoll P.D., Wolfgang M.J., Rosen C.L., D'Eustachio P.,
RA   Salzer J.L.;
RT   "Cloning of neurotrimin defines a new subfamily of differentially
RT   expressed neural cell adhesion molecules.";
RL   J. Neurosci. 15:2141-2156(1995).
CC   -!- FUNCTION: NEURAL CELL ADHESION MOLECULE.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- TISSUE SPECIFICITY: CENTRAL NERVOUS SYSTEM.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED AT HIGH LEVELS IN SEVERAL
CC   DEVELOPING PROJECTION SYSTEMS: IN NEURONS OF THE THALAMUS,
CC   SUBPLATE, AND LOWER CORTICAL LAMINAE IN THE FOREBRAIN AND IN THE
CC   PONTINE NUCLEUS, CEREBELLAR GRANULE CELLS, AND PURKINJE CELLS IN
CC   THE HINDBRAIN.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC   SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U16845; AAA67445.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00409; IG_like; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL         1     31       POTENTIAL.
FT   CHAIN         32    321       NEUROTRIMIN.
FT   PROPEP       322    344       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN        50    122       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN       150    208       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN       236    302       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID      57    115       POTENTIAL.
FT   DISULFID     157    201       POTENTIAL.
FT   DISULFID     243    295       POTENTIAL.
FT   CARBOHYD      44     44       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      70     70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     152    152       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     216    216       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     284    284       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     292    292       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     305    305       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     321    321       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID        321    321       GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   344 AA;  37998 MW;  CBB39BE53B33B224 CRC64;

Query Match             6.7%;
Best Local Similarity   24.9%;   Score 139.5;  DB 1;  Length 344;
Matches  66;  Conservative  40;  Mismatches 110;  Pred. No. 0.0009;  Indels  49;  Gaps  13;

Qy    3  RHLLTVPEAVGSGSGCNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLS--  59
Db   20  RLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDNRVTRVAWLNRSTILYAGN  79

Qy   60  ---VRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRL  109
Db   80  DKWCLDPRVVLLSN--TQTQY----SIEIQNVDYDEGPYTCSVQTDNHPKTSRV  128

Qy  110  HGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYFV  169
Db  129  H---LIVQVSPKIVEISSDISINEGNNISLTCIATGRPE-PTVTWR--HISPKAVGFV  180

Qy  170  PEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTIVIRCP-----QDTGGGI  224
Db  181  SEDEYLE----IQGITREQSGEYECSAS-NDVAAPVVRVRVNVTVNYPPYISEAKGTGPV  235

Qy  225  NIPGVL---SSLPSLGFSLPTWGK  245
Db  236  GQKGTLQCEASAVPSAEFQ---WFK  257

RESULT 3
G55A_CHICK
ID   G55A_CHICK   STANDARD;   PRT;   337 AA.
AC   Q9892;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neurite inhibitor GP55-A precursor (OBCAM protein gamma isoform).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
```

```
RA   Lodge A.P., Kim D. S., Howard M.R., McNamee C.J., Smith N., Moss D.J.;
RT   "Cloning of CEPU-s, a secreted isoform of CEPU-1, and OBCAM cDNAs from
RT   chick: structural diversity of IgLON family proteins.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 64-337 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=97157768; PubMed=9004047;
RA   Wilson D.J.A., Kim D. S., Clarke G.A., Marshall-Clarke S., Moss D.J.;
RT   "A family of glycoproteins (GP55) which inhibit neurite outgrowth,
RT   are members of the Ig superfamily and are related to OBCAM,
RT   neurotrimin, LAMP and CEPU-1.";
RL   J. Cell Sci. 109:3129-3138(1996).
CC   -!-  FUNCTION: INHIBITS NEURITE OUTGROWTH.
CC   -!-  SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!-  TISSUE SPECIFICITY: RESTRICTED TO THE NERVOUS SYSTEM.
CC   -!-  DEVELOPMENTAL STAGE: INCREASES DURING DEVELOPMENT FROM VERY LOW
CC        LEVELS AT EMBRYONIC DAY 10 AND IS MOST ABUNDANT AFTER HATCHING.
CC   -!-  SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC        SUBFAMILY.
CC   -!-  SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; Y08170; CAB41420.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL         1     20       BY SIMILARITY.
FT   CHAIN         21    317       NEURITE INHIBITOR GP55-A (POTENTIAL).
FT   PROPEP       318    337       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN        43    115       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN       143    201       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN       229    295       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID      50    108       POTENTIAL.
FT   DISULFID     150    194       POTENTIAL.
FT   DISULFID     236    288       POTENTIAL.
FT   CARBOHYD     133    133       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     277    277       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     285    285       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     298    298       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   337 AA;  36887 MW;  BAE717551856551E CRC64;

Query Match             6.4%;
Best Local Similarity   25.9%;   Score 133;  DB 1;  Length 337;
Matches  71;  Conservative  36;  Mismatches 107;  Pred. No. 0.0016;  Indels  60;  Gaps  16;

Qy    5  LLTVPEAVGSGSGCNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVL----  58
Db   16  LLFIP-GVPVRSGDATFPKAMDNVTVRQGESATLRCTVDDRVRRVAW-LNRSTILYAGND  73

Qy   59  --SVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRLH  110
Db   74  KWSIDNRVVILSN--TKTQY----SIKIHNVDYDEGPYTCSVQTDNHPKTSRVH  122

Qy  111  GSAYLTVQVMGELFIPSVNLVVAENEPCEVTCL-----PSHWTRLPDISWELGLLVSHSSY  166
Db  123  ----LIVQVPPQIVNISSDITVNEGSSVTLMCLAFGRPE----PTVTWR--HLSGKGQ  170

Qy  167  YFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTY-------IRCPQDTG  221
Db  171  GFVSEDEYLE----ITGITREQSGEYECSAV-NDAVPDVRVKVTVNYPPYISNAKNTG  225
```

```
Qy   222 GGINIPGVL----SSLPSLGFSLPTWGKVGLGIA 251
             |:|  |:||   ||
Db   226 ASVGQKGILQCEASAVPVAEFQ---WFKEDTRLA 256
```

RESULT 4

```
ID   DCC_MOUSE      STANDARD;      PRT;  1447 AA.
AC   P70211;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tumor suppressor protein DCC precursor.
GN   DCC.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS A, B AND C).
RC   STRAIN=BALB/c; TISSUE=Brain;
RA   Cooper H.M.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=96112625; PubMed=8570174;
RA   Cooper H.M., Armes P., Britto J., Gad J., Wilks A.F.;
RT   "Cloning of the mouse homologue of the deleted in colorectal cancer
RT   gene (mDCC) and its expression in the developing mouse embryo.";
RL   Oncogene 11:2243-2254(1995).
CC   -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 isoforms; A (shown here) and B; are
CC       produced by alternative splicing. A third isoform: C is
CC   -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED AT HIGH LEVELS IN THE
CC       DEVELOPING BRAIN AND NEURAL TUBE. IN ADULT, HIGHLY EXPRESSED IN
CC       BRAIN WITH VERY LOW LEVELS FOUND IN TESTIS, HEART AND THYMUS.
CC       ISOFORM C IS EXPRESSED ONLY IN THE EMBRYO.
CC   -!- DEVELOPMENTAL STAGE: LOW LEVELS IN EARLY GESTATION. HIGHEST LEVELS
CC       EXPRESSED DURING MID GESTATION. LEVELS DECREASE IN LATE GESTATION
CC       AND REMAIN AT THIS LEVEL IN THE ADULT.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE-III-LIKE DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; X85788; CAA59786.1; -.
DR   HSSP; P56276; 1TLK.
DR   MGD; MGI:94869; Dcc.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
KW   Glycoprotein; Transmembrane; Signal;
KW   Repeat; Anti-oncogene; Alternative initiation.
```

```
FT   SIGNAL        1     25
FT   CHAIN        26   1447       TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM A.
FT   CHAIN        85   1447       TUMOR SUPPRESSOR PROTEIN DCC, ISOFORM B.
FT   INIT_MET     85     85       FOR ISOFORM B.
FT   DOMAIN       26   1097       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1098   1122       POTENTIAL.
FT   DOMAIN     1123   1447       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       54    124       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      154    219       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      254    317       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      345    407       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      426    522       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      525    618       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      619    716       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      722    816       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      840    940       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      941   1042       FIBRONECTIN TYPE-III 6.
FT   DISULFID     61    117       BY SIMILARITY.
FT   DISULFID    161    212       BY SIMILARITY.
FT   DISULFID    261    310       BY SIMILARITY.
FT   DISULFID    352    400       BY SIMILARITY.
FT   CARBOHYD     60     60       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     94     94       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    318    318       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    478    478       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    628    628       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    702    702       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC    819    838       MISSING (IN ISOFORM C).
SQ   SEQUENCE   1447 AA;  158298 MW;  0D1F097C22D5B9F CRC64;
```

```
Qy     9 PEAVSGSGNEV-------------IEGPQNATVLKGSQARFNCTVSQGW--KLIMWAL 52
           :|:|: |:|               || : :| |  ||| |  |  :|
Db   217 PASIRGNEAEVRILSDPGLHRQLYFLQRPSNVIAIEGKDAVLECCVS-GYPPSFTWLR 275

Qy    53 SDMVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLH 110
           |:|: |:| | ||| ||| | || |  :|: || || |||
Db   276 GEEVI------------QLRSKKYSLLGG--SNLLISNVTDDDSGTYTCVAVTYKNENIS 320

Qy   111 GSAYLTVQVMGELFIPSVNLVAENEPCEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFV 169
           |||| || |:|   :|  |  |||    | ||     |  || ||  | : | |
Db   321 ASAELTVLVPPWFLNHPSNLYAYESMDIEFECAVS-GKPVPTVNMKNGDVVIPSDYFQI 379

Qy   170 PEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSA 207
           ||:    | ||   ||   || | :|  |  |  |:||
Db   380 VGGSNLR----ILGVVKSDEGFYQCVAENEAGNAQSSA 413
```

```
Query Match             6.3%;  Score 131;  DB 1;  Length 1447;
Best Local Similarity  25.2%;  Pred. No. 0.014;
Matches   55;  Conservative  33;  Mismatches  90;  Indels  40;  Gaps   9;
```

RESULT 5

```
ID   OPCM_HUMAN     STANDARD;      PRT;   345 AA.
AC   Q14982;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Opioid binding protein/cell adhesion molecule precursor (OBCAM)
DE   (Opioid-binding cell adhesion molecule) (OPCML).
GN   OPCML OR OBCAM.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Occipital cortex;
RX   MEDLINE=95237612; PubMed=7721093;
RA   Shark K.B., Lee N.M.;
RT   "Cloning, sequencing and localization to chromosome 11 of a cDNA
RT   encoding a human opioid-binding cell adhesion molecule (OBCAM).";
```

RESULT 6

RL   Gene 155:213-217(1995).
CC -!- FUNCTION: BINDS OPIOIDS IN THE PRESENCE OF ACIDIC LIPIDS; PROBABLY
CC   INVOLVED IN CELL CONTACT.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC   similarity).
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC   SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L34774; AAA36387.1; -.
DR   Genew; HGNC:8143; OPCML.
DR   MIM; 600632; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL   1   27
FT   CHAIN   28   322   OPIOID BINDING PROTEIN/CELL ADHESION
FT   MOLECULE.
FT   PROPEP   323   345   REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN   50   122   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN   150   209   IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN   237   303   IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID   57   115   POTENTIAL.
FT   DISULFID   157   202   POTENTIAL.
FT   DISULFID   244   296   POTENTIAL.
FT   CARBOHYD   44   44   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   70   70   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   140   140   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   285   285   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   293   293   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   306   306   N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID   322   322   GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   345 AA; 38007 MW; E7AD17BEA1AA3FF4 CRC64;

Query Match   6.3%;
Best Local Similarity   25.6%;
Matches   73; Conservative 34; Mismatches 108; Indels 70; Gaps 17;

Score 130; DB 1; Length 345;
Pred. No. 0.0028;

Qy   3 RHLLTVPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVL--- 58
Db   20 RLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAVW-LNRSTILYAG 78

Qy   59 ----SVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSR 108
Db   79 NDKWSIDPRVILVN--TPTQY------SIMIQNVDVYDEGPYTCSVQTDNHPKTSR   127

Qy   109 LHGSAVLTVQVWGELFIPSVNLVVAENEPCEVTCL---PSHWTLRLPDISWELGLLVSHS 164
Db   128 VH---LIVQVPPQIMNISSDITVNEGSSVTLLCLAIGRPE----PTVTWR----HL   172

Qy   165 SYY---FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV----IR   215
Db   173 SVKEGGFVSEDEYLE---ISDIKRDQSGEYECSAL-NDVAAPDVRKVKITVNYPPYIS 227

Qy   216 CPQDTGGGINIPGVL----SSLPSLGFSLPTWGK----VGLGLAG   252
Db   228 KAKNTGVSVGOKGILSCEASAVPMAEFQ--WFKEETRLATGLDG   269

ID   OPCM_BOVIN   STANDARD;   PRT;   345 AA.
AC   P11834;
DT   01-OCT-1989 (Rel. 12, Created)
DT   01-OCT-1989 (Rel. 12, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Opioid binding protein/cell adhesion molecule precursor (OB-CAM)
DE   (Opioid-binding protein/cell adhesion molecule) (OPCML).
GN   OPCML OR OBCAM OR OCAM.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=89251576; PubMed=2721489;
RA   Schofield P.R., McFarland K.C., Hayflick J.S., Wilcox J.N.,
RA   Cho T.M., Roy S., Lee N.M., Loh H.H., Seeburg P.H.;
RT   "Molecular characterization of a new immunoglobulin superfamily
RT   protein with potential roles in opioid binding and cell contact.";
RL   EMBO J. 8:489-495(1989).
CC -!- FUNCTION: BINDS OPIOIDS IN THE PRESENCE OF ACIDIC LIPIDS; PROBABLY
CC   INVOLVED IN CELL CONTACT.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor (By
CC   similarity).
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. IGLON
CC   SUBFAMILY.
CC -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; X12672; CAA31192.1; -.
DR   PIR; S03199; S03199.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal.
FT   SIGNAL   1   27
FT   CHAIN   28   322   OPIOID BINDING PROTEIN/CELL ADHESION
FT   MOLECULE.
FT   PROPEP   323   345   REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN   50   122   IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN   150   209   IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN   237   303   IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID   57   115   POTENTIAL.
FT   DISULFID   157   202   POTENTIAL.
FT   DISULFID   244   296   POTENTIAL.
FT   CARBOHYD   44   44   N-LINKED (GLCNAC...) (PROBABLE).
FT   CARBOHYD   70   70   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   140   140   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   285   285   N-LINKED (GLCNAC...) (PROBABLE).
FT   CARBOHYD   293   293   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   306   306   N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID   322   322   GPI-ANCHOR (POTENTIAL).
SQ   SEQUENCE   345 AA; 37914 MW; D1ECC8D9E7D8CB19 CRC64;

Query Match   6.2%;
Best Local Similarity   25.7%;
Matches   72; Conservative 33; Mismatches 109; Indels 66; Gaps 16;

Score 129; DB 1; Length 345;
Pred. No. 0.0034;

Qy   3 RHLLTVPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDNVVL--- 58

```
RESULT 7
ID   CEPU_CHICK    STANDARD;    PRT;    353 AA.
AC   Q90773;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   CEPU-1 protein precursor.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=96370549; PubMed=8774445;
RA   Spaltmann F., Bruemmendorf T.;
RT   "CEPU-1, a novel immunoglobulin superfamily molecule, is expressed by
RT   developing cerebellar Purkinje cells.";
RL   J. Neurosci. 16:1770-1779(1996).
CC   -!- FUNCTION: IT MAY BE A CELLULAR ADDRESS MOLECULE SPECIFIC TO
CC       PURKINJE CELLS. IT MAY REPRESENT A RECEPTOR OR A SUBUNIT OF A
CC       RECEPTOR COMPLEX.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A MAJOR FORM AND A MINOR FORM
CC       (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: FOUND ON THE DENDRITES, SOMATA AND AXONS OF
CC       DEVELOPING PURKINJE CELLS. UNDETECTABLE ON OTHER NEURONS LIKE
CC       GOLGI OR GRANULE CELLS.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED BY DEVELOPING CEREBELLAR PURKINJE
CC       CELLS. EXPRESSION COINCIDES WITH THE GROWTH OF THE DENDRITIC TREE,
CC       AFTER PURKINJE CELLS HAVE FINISHED THEIR MIGRATION FROM THE
CC       VENTRICULAR ZONE (FROM E15 UNTIL E21). EXPRESSED IN THE ADULT.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE IMUNOGLOBULIN SUPERFAMILY. IGLON
CC       SUBFAMILY.
CC   -!-
```

```
DR   EMBL; Z72497; CAA96578.1; -.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 2.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal; Alternative splicing.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    330       CEPU-1 PROTEIN.
FT   PROPEP      331    353       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       48    120       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      148    206       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      234    300       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID     55    113       POTENTIAL.
FT   DISULFID    155    199       POTENTIAL.
FT   DISULFID    241    293       POTENTIAL.
FT   CARBOHYD     42     42       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD     68     68       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    150    150       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    282    282       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    290    290       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   CARBOHYD    303    303       N-LINKED (GLCNAC...)  (POTENTIAL).
FT   LIPID       330    330       GPI-ANCHOR (POTENTIAL).
FT   VARSPLIC    310    330       MISSING (IN MAJOR ISOFORM).
SQ   SEQUENCE   353 AA;  38736 MW;  2550C48591EBBBA6 CRC64;

Query Match             6.2%;  Score 129;  DB 1;  Length 353;
Best Local Similarity  25.0%;  Pred. No. 0.0035;
Matches   66;  Conservative  38;  Mismatches  92;  Indels  68;  Gaps  14;

Qy     7  TVPEAVGSSGSGNEVIEGPQNATVLKGSQARFNCTVSQ------GW--KLIMWALSDMVVL   58
Db    35  TFPKAM------DNVTVRQGESATLRCSVDNRVTRVAWLNRSSILYAGNDKWCL         82

Qy    59  SVRPMEPIIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSRLHGS  112
Db    83  DPRVLLANTKTQYSIQ--------------------IHDVDYYDEGPYTCSVQTDNHPKTSRVH--  127

Qy   113  AYLTVQVMGELFIPSVNLVVAAENEPCEVTCLPSHWTRLPD--ISWELGLLVSHSSYYFVP  170
Db   127

Qy   128  --LIVQVSPKITETSSDISINEGGNVSLTCIA--TGRDPTIITWR--HISPKAVGFIS     179

Qy   171  EPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV------IRCPQDTGGGIN  225
Db   180  EDEYLE----ITGITREQSGEYECSAS-NDVAAPVVQRVKVTVNYPPYISDAKSTGVPVG   234

Qy   226  IPGVL----SSLPSLGFSLPTWGK                                      245
Db   235  QKGILMCEASAVPSADFQ---WYK                                      255
```

```
RESULT 8
ID   OPCM_RAT     STANDARD;    PRT;    345 AA.
AC   P32736; Q01654; P32735; Q01653;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Opioid binding protein/cell adhesion molecule precursor (OBCAM)
DE   (Opioid-binding cell adhesion molecule) (OPCML).
GN   OPCML OR OBCAM.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92347701; PubMed=1339369;
RA   Lippman D.A., Lee N.M., Loh H.H.;
RT   "Opioid-binding cell adhesion molecule (OBCAM)-related clones from a
RT   rat brain cDNA library.";
RL   Gene 117:249-254(1992).
RN   [2]
RP   SEQUENCE OF 195-214, AND GPI-ANCHOR.
RX   MEDLINE=95198094; PubMed=7891157;
RA   Struyk A.F., Canoll P.D., Wolfgang M.J., Rosen C.L., D'Eustachio P.,
RA   Salzer J.L.;
```

The alignment rows within the left/lower part (query/db pairs):

```
Qy    20  RLLFIVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAG   78
Qy    59  ---SVRPMEPIIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSR  108
Db    79  NDKWSIDPRVILVN---TPTQY-------SIMIQNVDYYDEGPYTCSVQTDNHPKTSR    127
Qy   109  LHGSAYLTVQVMGELFIPSVNLVVAAENEPCEVTCL-----PSHWTRLPDISWELGLLVSHS  164
Db   128  VH---LIVQVPQIMNISSDVTVNEGGSVTLLCLAIGRPE-----PTVTWR----HL      172
Qy   165  SYY----FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-----IR  215
Db   173  SVKEGQGFVSEDEYLE---ISDIKRDQSGEYECSAL-NDVAAPDVRKVKITVNYPPYIS   227
Qy   216  CPQDTGGGINIPGVL----SSLPSLGFSLPTWGKVGLGLA                     251
Db   228  KAKNTGVSVGQKGILSCEASAVPMAEFQ---WFKEDTRLA                     264
```

```
RT   "Cloning of neurotrimin defines a new subfamily of differentially
     expressed neural cell adhesion molecules.";
RL   J. Neurosci. 15:2141-2156(1995).
CC   -!- FUNCTION: BINDS OPIOIDS IN THE PRESENCE OF ACIDIC LIPIDS; PROBABLY
CC       INVOLVED IN CELL CONTACT.
CC   -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       MAY BE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS. IGLON
CC       SUBFAMILY.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M88710; AAA40859.1; -.
DR   EMBL; M88711; AAA40860.1; -.
DR   EMBL; M88709; AAA40858.1; -.
DR   PIR; JC1239; JC1239.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 3.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 3.
KW   Immunoglobulin domain; Cell adhesion; Glycoprotein; GPI-anchor;
KW   Repeat; Signal; Alternative splicing.
FT   SIGNAL        1     27
FT   CHAIN        28    322       OPIOID BINDING PROTEIN/CELL ADHESION
FT                                MOLECULE.
FT   PROPEP      323    345       REMOVED IN MATURE FORM (POTENTIAL).
FT   DOMAIN       50    122       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      150    209       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      237    303       IG-LIKE C2-TYPE DOMAIN 3.
FT   DISULFID     57    115       POTENTIAL.
FT   DISULFID    157    202       POTENTIAL.
FT   DISULFID    244    296       POTENTIAL.
FT   CARBOHYD     44     44       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     70     70       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    140    140       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    285    285       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    293    293       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    306    306       N-LINKED (GLCNAC...) (POTENTIAL).
FT   LIPID       322    322       GPI-ANCHOR (POTENTIAL).
FT   VARSPLIC      1     27       MGVCGYLFLPWKCLVVVSLRLLFLVPT -> MYHPAYWIVF
FT                                SATTALLFIP (IN ISOFORM 2).
SQ   SEQUENCE   345 AA;  38067 MW;  A3181B0753F96S8E CRC64;

Query Match             6.2%;  Score 128;  DB 1;  Length 345;
Best Local Similarity   25.3%;  Pred. No. 0.0041;
Matches  73;  Conservative  35;  Mismatches 110;  Indels  70;  Gaps  17;

Qy      3 RHLLTVPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVL--- 58
          |  || || |     ::|  :| | |  |  ::| |  ||: ||  :: |   ||
Db     20 RLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAG 78

Qy     59 ----SVRPMEPITINDRFTSQRYDQGGNFTSEMIHNVEPSDSGNIRCSLQ-----NSR 108
              :|:| :| :|||  :    ||: |  |  ::|  :  | |  |  |      |:|
Db     79 NDKWSIDPRVILLVN--TPTQY------SIMIQNVDVDEGPYTCSVQTDNHPKTSR 127

Qy    109 LHGSAVLTVQVMGELFIPSVNLVVAENEPCEVTCL-----PSHWTRLPDISWELGILVSHS 164
          | :| ||||:    |  |  |  |:  |  |  |      |  |   ||  |    |  ||
Db    128 VH---LIVQVPPQINISSDITVNISSVTLLCLAIGRPE-----PTVTWR-----HL 172

Qy    165 SYY----FVPEPSDLQQSAVSILALTFQQSNGTLTCVATWKSLKARKSATVNLTV--- 215
          : |     | |   :| :   :|| |  | | ||  |:|  |  |  |:   :|  :|
Db    128 VH---LIVQVPPQINISSDITVNISSVTLLCLAIGRPE-----PTVTWR-----HL 172

Qy    165 SYY----FVPEPSDLQQSAVSILALTFQQSNGTLTCVATWKSLKARKSATVNLTV---IR 215
          : |     | |   :| :   :|| |  | | ||  |:|  |  |  |:   :|  :|  :|
Db    173 SVKEGQGFVSEDEYLE----ISDIKRDQSGEYECSAL-NDVAAPDVRKVKITVNPPYIS 227
```

RESULT 9

```
ID   DCC_HUMAN      STANDARD;   PRT;   1447 AA.
AC   P43146;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tumor suppressor protein DCC precursor (Colorectal cancer suppressor).
GN   DCC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95011532; PubMed=7926722;
RA   Hedrick L., Cho K.R., Fearon E.R., Wu T.-C., Kinzler K.W.,
RA   Vogelstein B.;
RT   "The DCC gene product in cellular differentiation and colorectal
RT   tumorigenesis.";
RL   Genes Dev. 8:1174-1183(1994).
RN   [2]
RP   SEQUENCE OF 1-750 FROM N.A.
RX   MEDLINE=90100559; PubMed=2294591;
RA   Fearon E.R., Cho K.R., Nigro J.M., Kern S.E., Simons J.W.,
RA   Ruppert J.M., Hamilton S.R., Preisinger A.C., Thomas G., Kinzler K.W.,
RA   Vogelstein B.;
RT   "Identification of a chromosome 18q gene that is altered in
RT   colorectal cancers.";
RL   Science 247:49-56(1990).
RN   [3]
RP   SEQUENCE OF 107-472 FROM N.A. (SCRAMBLED EXONS).
RX   MEDLINE=91121517; PubMed=1991322;
RA   Cho K.R., Oliner J.D., Simons J.W., Hedrick L., Fearon E.R.,
RA   Preisinger A.C., Hedge P., Silverman G.A., Vogelstein B.;
RT   "The DCC gene: structural analysis and mutations in colorectal
RT   carcinomas.";
RL   Genomics 19:525-531(1994).
RN   [4]
RP   GENE STRUCTURE, AND VARIANTS CARCINOMA HIS-1375.
RX   MEDLINE=94245241; PubMed=8180295;
RA   Oliner J.D., Kinzler K.W., Vogelstein B.;
RL   Cell 64:607-613(1991).
RN   [5]
RP   VARIANT CARCINOMA THR-168, AND VARIANT GLY-201.
RX   MEDLINE=94243823; PubMed=8187090;
RA   Miyake S., Nagai K., Yoshino K., Oto M., Endo M., Yuasa Y.;
RT   "Point mutations and allelic deletion of tumor suppressor gene DCC in
RT   human esophageal squamous cell carcinomas and their relation to
RT   metastasis.";
RL   Cancer Res. 54:3007-3010(1994).
CC   -!- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR GENE.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: FOUND IN AXONS OF THE CENTRAL AND PERIPHERAL
CC       NERVOUS SYSTEM AND IN DIFFERENTIATED CELL TYPES OF THE INTESTINE.
CC   -!- DISEASE: COLORECTAL TUMORS THAT LOST THEIR CAPACITY TO
CC       DIFFERENTIATE INTO MUCUS PRODUCING CELLS UNIFORMLY LACK DCC
CC       EXPRESSION. INACTIVATION OF DCC DUE TO ALLELIC DELETION AND/OR
CC       POINT MUTATIONS MAY CAUSE BOTH LYMPHATIC AND HEMATOGENOUS
CC       METASTASIS OF OESOPHAGEAL SQUAMOUS CELL CARCINOMAS.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

Qy   216 CPQDTGGGINIPGVL----SSLPSLGFSLPTWGK----VGLLAGTML 255
          : || | | :::| :| | | | ;
Db   228 KAKNTGVSVGQKGILSCEASAVPMAEFQ---WFKEDTRLATGLLDGVRI 272

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X76132; CAA53735.1; -.
DR   EMBL; M32292; AAA35751.1; -.
DR   EMBL; M32286; AAA52174.1; -.
DR   EMBL; M32288; AAA52175.1; ALT_SEQ.
DR   EMBL; M32290; AAA52176.1; -.
DR   EMBL; M63696; AAA52177.1; -.
DR   EMBL; M63700; AAA52178.1; -.
DR   EMBL; M63702; AAA52179.1; -.
DR   EMBL; M63718; AAA52180.1; -.
DR   EMBL; M63698; AAA52181.1; -.
DR   PIR; A54100; A54100.
DR   PIR; A40098; A40098.
DR   PIR; A38442; A38442.
DR   HSSP; P56276; 1TLK.
DR   Genew; HGNC:2701; DCC.
     MIM; 120470; -.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003362; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IGc2; 3.
KW   Glycoprotein; Immunoglobulin domain; Transmembrane; Signal;
KW   Repeat; Anti-oncogene; Disease mutation; Polymorphism.
FT   SIGNAL        1     25
FT   CHAIN        26   1447
FT   DOMAIN       26   1097  EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1098   1122  POTENTIAL.
FT   DOMAIN     1123   1447  CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       54    124  IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      154    219  IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      254    317  IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      345    407  IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      426    522  FIBRONECTIN TYPE-III 1.
FT   DOMAIN      525    618  FIBRONECTIN TYPE-III 2.
FT   DOMAIN      619    716  FIBRONECTIN TYPE-III 3.
FT   DOMAIN      722    816  FIBRONECTIN TYPE-III 4.
FT   DOMAIN      840    940  FIBRONECTIN TYPE-III 5.
FT   DOMAIN      941   1042  FIBRONECTIN TYPE-III 6.
FT   DISULFID     61    117  BY SIMILARITY.
FT   DISULFID    161    212  BY SIMILARITY.
FT   DISULFID    261    310  BY SIMILARITY.
FT   DISULFID    352    400  BY SIMILARITY.
FT   CARBOHYD     94     94  N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    299    299  N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    318    318  N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    478    478  N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    628    628  N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    702    702  N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    168    168  N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARIANT     201    201  M -> T (IN OESOPHAGEAL CARCINOMA).
FT                           /FTId=VAR_003909.
FT   VARIANT    1375   1375  R -> G.
FT                           /FTId=VAR_003910.
FT   VARIANT     201    201  P -> H (IN A COLORECTAL CARCINOMA).
FT                           /FTId=VAR_003911.
FT   CONFLICT    138    138  MISSING (IN REF. 3).
FT   CONFLICT    233    329  MISSING (IN REF. 3).
FT   CONFLICT    421    421  MISSING (IN REF. 3).
SQ   SEQUENCE   1447 AA;  158456 MW;  4A8612766ED0471F CRC64;

Query Match             6.2%;  Score 128;  DB 1;  Length 1447;
```

```
Best Local Similarity   24.8%;
Matches   60;  Conservative   33;  Mismatches   99;  Indels   50;  Gaps   10;

RESULT 10
KMLS_HUMAN

ID   KMLS_HUMAN     STANDARD;      PRT;   1914 AA.
AC   Q15746; Q14844; Q16794; O95796; O95797; O95798; O95799; Q9UBG5;
AC   Q9UIT9;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Myosin light chain kinase, smooth muscle and non-muscle isozymes
DE   (EC 2.7.1.117) (MLCK) [Contains: Telokin (Kinase related protein)
DE   (KRP)].
GN   MYLK OR MLCK.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Umbilical vein endothelial cells;
RX   MEDLINE=97304466; PubMed=9160829;
RA   Garcia J.G.N., Lazar V.L., Gilbert-Mcclain L.I., Gallagher P.J.,
RA   Verin A.D.;
RT   "Myosin light chain kinase in endothelium: molecular cloning and
RT   regulation.";
RL   Am. J. Respir. Cell Mol. Biol. 16:489-494(1997).
RN   [2]
RP   REVISIONS.
RA   Birukov K.G., Garcia J.G.N.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 2; 3A; 3B AND 4).
RC   TISSUE=Umbilical vein;
RX   MEDLINE=99216419; PubMed=10198165;
RA   Lazar V.L., Garcia J.G.N.;
RT   "A single human myosin light chain kinase gene (MLCK; MYLK).";
RL   Genomics 57:256-267(1999).
RN   [4]
RP   REVISIONS (ISOFORM 2).
RA   Birukov K.G., Garcia J.G.N.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 923-1914 FROM N.A.
RC   TISSUE=Hippocampus;
RX   MEDLINE=96121365; PubMed=8575746;
RA   Potier M.-C., Chelot E., Pekarsky Y., Gardiner K., Rossier J.,
RA   Turnell W.G.;
RT   "The human myosin light chain kinase (MLCK) from hippocampus:
RT   cloning, sequencing, expression, and localization to 3qcen-q21.";
RL   Genomics 29:562-570(1995).

Qy   14 SGSGNEV-----------------------------IEGPQNATVLKGSQARRNCTVSQGW--KLIMNALSDM 55
         : |||                             ||| : :|| :::| |  |:    || :  |:  :|
Db  220 SRTGNEAEVRILSDPGLHROLIFLQRPSNVVAIEGKDAVLECCVS--GYPPPSFTWLRGEE 278

Qy   56 VVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSA 113
         :| : : :| ||  | :  |     || | | :||  : |||   ||    |   |  ||
Db  279 VI--------QLRSKRYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNENISASA 323

Qy  114 YLTVQVWGELFIPSVNLVVAENEPCEVTLCLPSHWTRLPDISW-ELGLLVSHSSSYFVPEP 172
         : |: : | | :| ||  |    | ||  | |   || ||::   | ||:: |:| :| |
Db  324 ELTVLVPPWFLNHPSNLYAYESMDIEFECTVS-GKPVPTVNWMKNGDVVIPSDYFQIVGG 382

Qy  173 SDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGINIPGVLSS 232
         |:|    | | || | || | |||  |  |  |||  ::|:|     :  |  ||:|||
Db  383 SNLR------ILCGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS-------VLPS 430

Qy  233 LP 234
         | |
Db  431 AP 432
```

```
RN   [6]
RP   SEQUENCE OF 1614-1914 FROM N.A.
RC   TISSUE=Lung, and Placenta;
RX   MEDLINE=20007838; PubMed=10536370;
RA   Watterson D.M., Schavocky J.P., Guo L., Weiss C., Chlenski A.,
RA   Shirinsky V.P., Van Eldik L.J., Haiech J.;
RT   "Analysis of the kinase-related protein gene found at human chromosome
RT   3q21 in a multi-gene cluster: organization, expression, alternative
RT   splicing and polymorphic marker.";
RL   J. Cell. Biochem. 75:481-491(1999).
RN   [7]
RP   SEQUENCE OF 1456-1914 FROM N.A.
RA   Watterson M.D.;
RL   Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.
RC   TISSUE=Placenta;
CC   -!- FUNCTION: CALCIUM/CALMODULIN-DEPENDENT ENZYME RESPONSIBLE FOR
CC       SMOOTH MUSCLE CONTRACTION VIA PHOSPHORYLATION OF A SPECIFIC SERINE
CC       IN THE N-TERMINUS OF MYOSIN LIGHT CHAINS (MLC), AN EVENT THAT
CC       FACILITATES MYOSIN INTERACTION WITH ACTIN FILAMENTS. CENTRAL
CC       DETERMINANT IN THE DEVELOPMENT OF VASCULAR PERMEABILITY AND TISSUE
CC       EDEMA FORMATION. IN THE NERVOUS SYSTEM IT HAS BEEN SHOWN TO
CC       CONTROL THE GROWTH INITIATION OF ASTROCYTIC PROCESSES IN CULTURE
CC       AND TO PARTICIPATE IN TRANSMITTER RELEASE AT SYNAPSES FORMED
CC       BETWEEN CULTURED SYMPATHETIC GANGLION CELLS. CRITICAL PARTICIPANT
CC       IN SIGNALING SEQUENCES THAT RESULT IN FIBROBLAST APOPTOSIS.
CC   -!- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin
CC       light-chain] phosphate.
CC   -!- SUBUNIT: TELOKIN BINDS CALMODULIN.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; A NON-MUSCLE FORM (THE
CC       LONGEST FORM), A SMOOTH-MUSCLE FORM AND TELOKIN (A C-TERMINAL
CC       SECTION WITH NO CATALYTIC ACTIVITY); ARE PRODUCED BY ALTERNATIVE
CC       INITIATION.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 6 ISOFORMS; 1 (SHOWN HERE), 2, 3A,
CC       3B, 4 AND DEL-1790; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: SMOOTH MUSCLE AND NON-MUSCLE ISOZYMES ARE
CC       EXPRESSED IN A WIDE VARIETY OF ADULT AND FETAL TISSUES AND IN
CC       CULTURED ENDOTHELIUM WITH QUALITATIVE EXPRESSION APPEARING TO BE
CC       NEITHER TISSUE- NOR DEVELOPMENT-SPECIFIC. NON-MUSCLE ISOFORM 2 IS
CC       THE DOMINANT SPLICE VARIANT EXPRESSED IN VARIOUS TISSUES. TELOKIN
CC       HAS BEEN FOUND IN A WIDE VARIETY OF ADULT AND FETAL TISSUES.
CC   -!- PTM: MLCK IS PROBABLY DOWN-REGULATED BY PHOSPHORYLATION.
CC   -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.
```

```
DR   EMBL; U48959; AAC18423.2; -.
DR   EMBL; AF069601; AAD15921.2; -.
DR   EMBL; AF069602; AAD15922.1; -.
DR   EMBL; AF069603; AAD15923.1; -.
DR   EMBL; AF069604; AAD15924.1; -.
DR   EMBL; X85337; CAA59685.1; -.
DR   EMBL; AF096771; AAD51380.1; -.
DR   EMBL; AF096766; AAD51380.1; JOINED.
DR   EMBL; AF096767; AAD51380.1; JOINED.
DR   EMBL; AF096768; AAD51380.1; JOINED.
DR   EMBL; AF096769; AAD51380.1; JOINED.
DR   EMBL; AF096770; AAD51380.1; JOINED.
DR   EMBL; AF096774; AAD54018.1; -.
DR   EMBL; AF096771; AAD51381.1; -.
DR   EMBL; AF096769; AAD51381.1; JOINED.
DR   EMBL; AF096770; AAD51381.1; JOINED.
DR   HSSP; Q63450; 1A06.
DR   Genew; HGNC:7590; MYLK.
DR   MIM; 600922; -.

DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR022290; Ser_thr_pkinase.
DR   Pfam; PF00041; fn3; 1.
DR   Pfam; PF00069; Pkinase; 1.
DR   Pfam; PF00047; ig; 8.
DR   PRINTS; PR00014; FNTYPEIII.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM000060; FN3; 1.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGc2; 8.
DR   SMART; SM000220; S_TKc; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ST; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ATP; 1.
KW   ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;
KW   Transferase; Serine/threonine-protein kinase; Calmodulin-binding;
KW   Alternative initiation; Alternative splicing.
FT   INIT_MET   1761   1761
FT   CHAIN       923   1914   MYOSIN LIGHT CHAIN KINASE, SMOOTH-MUSCLE
FT                            ISOZYME.
FT   CHAIN      1761   1914   TELOKIN.
FT   CHAIN         1   1914   MYOSIN LIGHT CHAIN KINASE, NON-MUSCLE
FT                            ISOZYME.
FT   DOMAIN     1343   1413   FOR MYOSIN LIGHT CHAIN KINASE, SMOOTH-
FT                            MUSCLE ISOZYME.
FT   DOMAIN     1464   1719   FOR TELOKIN.
FT   DOMAIN     1711   1774   FIBRONECTIN TYPE-III.
FT   DOMAIN     1824   1891   PROTEIN KINASE.
FT   NP_BIND    1470   1478   CALMODULIN-BINDING.
FT   BINDING    1493   1493   IG-LIKE C2-TYPE DOMAIN.
FT   ACT_SITE   1585   1585   ATP (BY SIMILARITY).
FT   DOMAIN     1906   1914   ATP (BY SIMILARITY).
FT                            BY SIMILARITY.
FT   DOMAIN      868    895   POLY-GLU.
FT   REPEAT      868    895   5 X 28 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT      896    923   I-1.
FT   REPEAT      924    951   I-2.
FT   REPEAT      952    979   I-3.
FT   REPEAT      980    998   I-4.
FT   DOMAIN      999   1063   I-5.
FT   REPEAT      999   1003   6 X 12 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT     1004   1015   II-1. (INCOMPLETE).
FT   REPEAT     1016   1027   II-2.
FT   REPEAT     1028   1039   II-3.
FT   REPEAT     1040   1051   II-4.
FT   REPEAT     1052   1063   II-5.
FT                            II-6.
FT   VARSPLIC    437    506   (INCOMPLETE).
FT                            VSGIPKPEVAWFLEGTPVRRQEGSIEVYEDAGSHYICLLKA
FT                            RTRDSGTYSCTASNAQGQVSCSWTLQVER -> G (IN
FT                            ISOFORM 2 AND ISOFORM 3B).
FT   VARSPLIC   1433   1439   DEVEVSD -> MKWRCQT (IN ISOFORM 3A,
FT                            ISOFORM 3B AND ISOFORM 4).
FT   VARSPLIC   1473   1546   GKFGGQVFRLVEKKTPRKVWAGKFFKAYSAKEKENIRQEISIM
FT                            NCLHHPKLVQCVDAFEEKANIVMVLEIVSGGEL -> L
FT                            (IN ISOFORM 4).
FT   VARSPLIC   1655   1705   MISSING (IN ISOFORM 3A AND ISOFORM 3B).
FT   VARSPLIC   1790   1790   MISSING (IN ISOFORM DEL-1790).
FT   CONFLICT    933    933   V -> M (IN REF. 5).
FT   CONFLICT    963    963   S -> P (IN REF. 5).    AAD15922).
FT   CONFLICT   1022   1022   P -> A (IN REF. 5).
FT   CONFLICT   1048   1050   KPM -> EAH (IN REF. 5).
FT   CONFLICT   1162   1162   P -> L (IN REF. 3).    AAD15922/AAD15923).
FT   CONFLICT   1210   1210   L -> P (IN REF. 3).    AAD15923).
FT   CONFLICT   1280   1280   E -> D (IN REF. 3).    AAD15922/AAD15923).
FT   CONFLICT   1284   1284   M -> I (IN REF. 3).    AAD15922/AAD15923/
FT   CONFLICT   1300   1300   A -> G (IN REF. 5).    AAD15924).
FT   CONFLICT   1316   1316   L -> S (IN REF. 5).
FT   CONFLICT   1326   1326   T -> S (IN REF. 5).
```

```
FT   CONFLICT     1478   1478   V -> C (IN REF. 5).
FT   CONFLICT     1511   1511   S -> T (IN REF. 3;  AAD15922/AAD15923).
FT   CONFLICT     1563   1563   I -> T (IN REF. 5).
FT   CONFLICT     1609   1609   A -> P (IN REF. 5).
FT   CONFLICT     1609   1609   A -> V (IN REF. 5).
FT   CONFLICT     1639   1639   G -> R (IN REF. 5).
FT   CONFLICT     1639   1639   GY -> D (IN REF. 3;  AAD15922/AAD15923/
FT   CONFLICT     1640   1640   AAD15924).
```

```
Query Match                    6.2%;    Score 128;  DB 1;  Length 1914;
Best Local Similarity   20.3%;  Pred. No. 0.034;
Matches   86;  Conservative   58;  Mismatches 149;  Indels 130;  Gaps  19;

Qy     24  PQNATVLKGSQARFNCTVSQGW--KLIMWALSDMVVLSVRPMEPIITINDRFTSQRYDQGG   81
               ::|: :|: | |:| :| ::| :    :|:: :|::|  ||:| | |  :|:  ::|
Db     39  PRNLCIKEGATAPKFEGRV--RGYPEPQVTWH-------------------RNGQPITSGGRFL---LDCGI   85

Qy     82  NFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQV-----------------MGEL  123
               :: ||            :|||               :||:               :|:
Db     86  RGTFSLVIHAVHEEDRGKYTCEATNG--SGARQVIVELTVEGSFAKQLGQPVVSKTLGDR  143

Qy    124  FI-----------PSV----------PCEVTCLPSHWIRLPDISWELG-  158
            ||           |||          || :     |:|| |: : :|
Db    144  FSASAVETRPSIWGECPPKFATKLGRVVVKEGQMGRFSCKITGRPQ------PQVTWLKGN  198

Qy    159  LLVSHSSYFYFVPEPSDLQSAVSILALTPQSNGTLQTCVATWKSLKARKSA-------SGYNS  322
            |: | :| :|   |:  :  :|  : | :|| ||    | |  || | |       ::
Db    199  VPLQPSARVSVSEKNGMQ--VLEIHGVNQDDVGVYTCLVVNGSGKASMSAELSIQGLDSAN  257

Qy    211  LTVIRCPQDTGGGI-------NIPGVLSSLPSLGFSLPTWGKVGLGLGLAGTMLLTPTCLTIR  265
               :| |  |:|           |:    |::  ::|   |||:|:|:|  :|:|:: ||:|:|
Db    258  RSFVRETKATNSDVRKEVTNVISKESKLDSL--                              288

Qy    266  CCCCRRRCCGCNCCCRCFCCRRKRGFRIQFQKKSEKEKINKETETESGNEN---SGYNS  322
            |||| | |||| |||||||||| | |:|  | | ::||  :| | | ||:|:
Db    289  ------EAAAKSKNCSSPQRGGSPPWAANSQPQP-RESKLESCKDSPRTAPQTP  336

Qy    323  DEQKTTETASLPPKSCESSDPEQRNSSCGPPHQRADQR-----PPRPASHPQASFNIASPE  378
                   :|: ||  :   :|||:|                 ||||||||:  :||: |:
Db    337  VLQKTSSSITLQAARVQ---PEPRAPGLGVLSPSGEERKRPAPPRPATFPTRQPGLGSQD  393

Qy    379  KVS  381
            |||
Db    394  VVS  396
```

```
RESULT 11
ID   PTK7_HUMAN         STANDARD;    PRT;   1070 AA.
AC   Q13308; Q13417;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Tyrosine-protein kinase-like 7 precursor (Colon carcinoma kinase-4)
DE   (CCK-4).
GN   PTK7 OR CCK4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Colon carcinoma, and Placenta;
RX   MEDLINE=96074849; PubMed=7478540;
RA   Mossie K., Jallal B., Alves F., Sures I., Plowman G.D., Ullrich A.;
RT   "Colon carcinoma kinase-4 defines a new subclass of the receptor
RT   tyrosine kinase family.";
RL   Oncogene 11:2179-2184(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fibroblast;
RX   MEDLINE=97037064; PubMed=8882711;
RA   Park S.-K., Lee H.-S., Lee S.-T.;
RT   "Characterization of the human full-length PTK7 cDNA encoding a
RT   receptor protein tyrosine kinase-like molecule closely related to
RT   chick KLG.";
RL   J. Biochem. 119:235-239(1996).
RN   [3]
RP   SEQUENCE FROM N.A., AND REVISION TO 834.
RA   Lee S.-T., Park S.-K., Lee H.-S., Ji A.R., Jung J.W.;
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY FUNCTION AS A CELL ADHESION MOLECULE. LACKS PROBABLY
CC       THE CATALYTIC ACTIVITY OF TYROSINE KINASE. MAY BE CONNECTED TO THE
CC       PATHOPHYSIOLOGY OF COLON CARCINOMAS AND/OR MAY REPRESENT A TUMOR
CC       PROGRESSION MARKER.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LUNG, LIVER, PANCREAS,
CC       KIDNEY, PLACENTA AND MELANOCYTES. WEAKLY EXPRESSED IN THYROID
CC       GLAND, OVARY, BRAIN, HEART AND SKELETAL MUSCLE. ALSO EXPRESSED IN
CC       ERYTHROLEUKEMIA CELLS. BUT NOT EXPRESSED IN COLON.
CC   -!- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC       RECEPTOR SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC  -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC  -----------------------------------------------------------------------
DR   EMBL; U33635; AAA87565.1; -.
DR   EMBL; U40271; AAC50484.2; -.
DR   EMBL; AF447176; AAL39062.1; -.
DR   EMBL; AF447157; AAL39062.1; JOINED.
DR   EMBL; AF447158; AAL39062.1; JOINED.
DR   EMBL; AF447162; AAL39062.1; JOINED.
DR   EMBL; AF447164; AAL39062.1; JOINED.
DR   EMBL; AF447167; AAL39062.1; JOINED.
DR   EMBL; AF447170; AAL39062.1; JOINED.
DR   EMBL; AF447171; AAL39062.1; JOINED.
DR   EMBL; AF447173; AAL39062.1; JOINED.
DR   EMBL; AF447174; AAL39062.1; JOINED.
DR   EMBL; AF447175; AAL39062.1; JOINED.
DR   HSSP; P08631; 1AD5.
DR   Genew; HGNC:9618; PTK7.
DR   MIM; 601890;
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR002011; RTKinaseII.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 7.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Euk_pkinase; 1.
DR   SMART; SM00410; Euk_pkinase; 1.
DR   SMART; SM00408; IGc2; 5.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00239; RECEPTOR_TYR_KIN_II; FALSE_NEG.
DR   PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
KW   Receptor; Transmembrane; Signal; Glycoprotein; Cell adhesion;
KW   Immunoglobulin domain; Repeat.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1070       TYROSINE-PROTEIN KINASE-LIKE 7.
FT   DOMAIN       31    704       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    705    725       POTENTIAL.
FT   DOMAIN      726   1070       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       46    108       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      143    208       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      239    308       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      336    398       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      426    488       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      517    577       IG-LIKE C2-TYPE DOMAIN 6.
```

```
FT   DOMAIN        606    671       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN        796   1066       PROTEIN KINASE; INACTIVE.
FT   DISULFID       53    101       BY SIMILARITY.
FT   DISULFID      150    200       BY SIMILARITY.
FT   DISULFID      246    301       BY SIMILARITY.
FT   DISULFID      343    391       BY SIMILARITY.
FT   DISULFID      433    481       BY SIMILARITY.
FT   DISULFID      524    570       BY SIMILARITY.
FT   DISULFID      613    664       BY SIMILARITY.
FT   CARBOHYD      116    116       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      175    175       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      184    184       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      214    214       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      268    268       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      283    283       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      405    405       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      463    463       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      567    567       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD      646    646       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD       92     92       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT      147    147       P -> R (IN REF. 2 AND 3).
FT   CONFLICT      207    207       K -> T (IN REF. 2 AND 3).
FT   CONFLICT      495    496       S -> G (IN REF. 2 AND 3).
FT   CONFLICT      515    515       VL -> RV (IN REF. 2 AND 3).
FT   CONFLICT      881    881       G -> E (IN REF. 2 AND 3).
FT   CONFLICT      969    969       A -> P (IN REF. 2 AND 3).
FT   CONFLICT      992    992       S -> F (IN REF. 2 AND 3).
SQ   SEQUENCE   1070 AA;  118260 MW;  47CDF25B8E3698A5 CRC64;
```

```
Query Match             6.0%;   Score 125;   DB 1;   Length 1070;
Best Local Similarity  24.8%;   Pred. No. 0.029;
Matches   59;   Conservative   34;   Mismatches 109;   Indels  36;   Gaps  10;

Qy    11  AVGSGSGNEVIEGPQNATVLKGSQARFNCTVS-QGWKLIMWALSDMVVLSVRPMEPIITN   69
Db   218  SIADESFARVVLAPQDVVVARYEEAMHCQFSAQPPSLQWLFEDETPITNRSRPPHLRR     277

Qy    70  DRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSR-----LHGSAYLIVQVMGELF   124
Db   278  ATVFA------------NGSLLTQVRPRNAGIYRCIGQGQRGPPIILEATLHLAEIEDMPLF 328

Qy   125  IPSVNLVVAENEPCEVCLPSHWTRLPDISWE-LGL-LVSHSSYYFVPEPSDLQSAVSIL   182
Db   329  EPRVFTAGSEE------RVTCLPPKGLPEPSVWWEHAGVRLPTHGRVY------QKGHELVL 378

Qy   183  ALTPQSN-GTLTCVATWKSLKARKSATVNLIV------IRCPQDTGGGINIPGVLSSL    233
Db   379  ANIAESDAGVYTCHAA--NLAQQRRQDVNITVATVPSWLKKPQDSQLEEGKPGYLDCL    434
```

```
RESULT 12

ID   PGBM_HUMAN     STANDARD;      PRT;   4393 AA.
AC   P98160; Q16287;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Basement membrane-specific heparan sulfate proteoglycan core
DE   protein precursor (HSPG) (Perlecan) (PLC).
GN   HSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92129994; PubMed=1730768;
RA   Kallunki P., Tryggvason K.;
RT   "Human basement membrane heparan sulfate proteoglycan core protein: a
RT   467-kD protein containing multiple domains resembling elements of the
RT   low density lipoprotein receptor, laminin, neural cell adhesion
RT   molecules, and epidermal growth factor.";
RL   J. Cell Biol. 116:559-571(1992).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin, and Colon;
RX   MEDLINE=92235084; PubMed=1569102;
RA   Murdoch A.D., Dodge G.R., Cohen I., Tuan R.S., Iozzo R.V.;
RT   "Primary structure of the human heparan sulfate proteoglycan from
RT   basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT   domains homologous to the low density lipoprotein receptor, laminin,
RT   neural cell adhesion molecules, and epidermal growth factor.";
RL   J. Biol. Chem. 267:8544-8557(1992).
RN   [3]
RP   SEQUENCE OF 1018-1472 FROM N.A.
RC   TISSUE=Colon;
RX   MEDLINE=91365376; PubMed=1679749;
RA   Dodge G.R., Kovalszky I., Chu M.L., Hassell J.R., McBride O.W.,
RA   Yi H.F., Iozzo R.V.;
RT   "Heparan sulfate proteoglycan of human colon: partial molecular
RT   cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT   short arm of human chromosome 1.";
RL   Genomics 10:673-680(1991).
RN   [4]
RP   SEQUENCE OF 892-1398 FROM N.A.
RC   TISSUE=Fibrosarcoma;
RX   MEDLINE=94052171; PubMed=8234307;
RA   Cohen I.R., Graessel S., Murdoch A.D., Iozzo R.V.;
RT   "Structural characterization of the complete human perlecan gene and
RT   its promoter.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:10404-10408(1993).
RN   [5]
RP   SEQUENCE OF 1-21 FROM N.A.
RX   MEDLINE=92120660; PubMed=1685141;
RA   Kallunki P., Eddy R.L., Byers M.G., Kestila M., Shows T.B.,
RA   Tryggvason K.;
RT   "Cloning of human heparan sulfate proteoglycan core protein,
RT   assignment of the gene (HSPG2) to 1p36.1->p35 and identification of
RT   a BamHI restriction fragment length polymorphism.";
RL   Genomics 11:389-396(1991).
```

```
CC   -!- FUNCTION: This protein is an integral component of basement
CC       membranes. It is responsible for the fixed negative electrostatic
CC       charge and is involved in the charge-selective ultrafiltration
CC       properties. It serves as an attachment substrate for cells.
CC   -!- SUBUNIT: Purified perlecan has a strong tendency to aggregate in
CC       dimers or stellate structures. It interacts with other basement
CC       membrane components such as laminin, prolargin and collagen type
CC       IV.
CC   -!- SUBCELLULAR LOCATION: Extracellular.
CC   -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
CC   -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC       AND O-LINKED OLIGOSACCHARIDES.
CC   -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC   -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC   -!- SIMILARITY: CONTAINS 22 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LAMININ G-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
```

```
DR   EMBL; X62515; CAA44373.1; -.
DR   EMBL; M85289; AAA52700.1; -.
DR   EMBL; M64283; AAA52699.1; -.
DR   EMBL; S76436; AAB21121.2; -.
DR   EMBL; L22078; -; NOT_ANNOTATED_CDS.
DR   HSSP; P00740; 1EDM.
DR   Siena-2DPAGE; P98160; -.
DR   Genew; HGNC:5273; HSPG2.
```

```
DR   MIM; 142461; -.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR002172; LDL_recept_A.
DR   InterPro; IPR002034; Laminin_B.
DR   InterPro; IPR002049; Laminin_EGF.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR000082; SEA_domain.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00047; Ig; 22.
DR   Pfam; PF00052; laminin_B; 3.
DR   Pfam; PF00053; laminin_EGF; 7.
DR   Pfam; PF00054; laminin_G; 8.
DR   Pfam; PF00057; ldl_recept_a; 4.
DR   Pfam; PF01390; SEA; 1.
DR   PRINTS; PR00010; EGFBLOOD.
DR   ProDom; PD003031; Laminin_B; 3.
DR   SMART; SM00192; LDLa; 4.
DR   SMART; SM00181; LamB; 3.
DR   SMART; SM00282; LamG; 3.
DR   SMART; SM00180; EGF_Lam; 6.
DR   SMART; SM00001; EGF_like; 8.
DR   SMART; SM00408; IGc2; 22.
DR   PROSITE; PS00022; EGF_1; 9.
DR   PROSITE; PS01186; EGF_2; 5.
DR   PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
DR   PROSITE; PS50025; LAM_G_DOMAIN; 3.
DR   PROSITE; PS01209; LDLRA_1; 4.
DR   PROSITE; PS50068; LDLRA_2; 4.
DR   PROSITE; PS50024; SEA; 1.
KW   Signal; Basement membrane; Proteoglycan; Glycoprotein;
KW   Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;
KW   Extracellular matrix; Laminin EGF-like domain.
FT   SIGNAL     1     21       POTENTIAL.
FT   CHAIN     22   4393       BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT                             SULFATE PROTEOGLYCAN CORE PROTEIN.
FT   DOMAIN    80    194       SEA.
FT   DOMAIN   197    236       LDL-RECEPTOR CLASS A 1.
FT   DOMAIN   283    321       LDL-RECEPTOR CLASS A 2.
FT   DOMAIN   323    361       LDL-RECEPTOR CLASS A 3.
FT   DOMAIN   366    405       LDL-RECEPTOR CLASS A 4.
FT   DOMAIN   405    506       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN   523    532       LAMININ EGF-LIKE 1 (N-TERMINAL).
FT   DOMAIN   533    732       LAMININ DOMAIN IV 1 (DOMAIN III A).
FT   DOMAIN   733    765       LAMININ EGF-LIKE 1 (C-TERMINAL).
FT   DOMAIN   766    815       LAMININ EGF-LIKE 2.
FT   DOMAIN   816    873       LAMININ EGF-LIKE 3.
FT   DOMAIN   881    925       LAMININ EGF-LIKE 4 (INCOMPLETE).
FT   DOMAIN   926    935       LAMININ EGF-LIKE 5 (N-TERMINAL).
FT   DOMAIN   936   1127       LAMININ DOMAIN IV 2 (DOMAIN III B).
FT   DOMAIN  1128   1160       LAMININ EGF-LIKE 5 (C-TERMINAL).
FT   DOMAIN  1161   1210       LAMININ EGF-LIKE 6.
FT   DOMAIN  1211   1267       LAMININ EGF-LIKE 7.
FT   DOMAIN  1277   1326       LAMININ EGF-LIKE 8.
FT   DOMAIN  1327   1336       LAMININ EGF-LIKE 9 (N-TERMINAL).
FT   DOMAIN  1337   1531       LAMININ DOMAIN IV 3 (DOMAIN III C).
FT   DOMAIN  1532   1564       LAMININ EGF-LIKE 9 (C-TERMINAL).
FT   DOMAIN  1565   1614       LAMININ EGF-LIKE 10.
FT   DOMAIN  1615   1672       LAMININ EGF-LIKE 11.
FT   DOMAIN  1679   1773       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN  1774   1867       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN  1868   1957       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN  1958   2053       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN  2054   2153       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN  2154   2246       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN  2247   2342       IG-LIKE C2-TYPE DOMAIN 8.
FT   DOMAIN  2343   2438       IG-LIKE C2-TYPE DOMAIN 9.
FT   DOMAIN  2439   2535       IG-LIKE C2-TYPE DOMAIN 10.
FT   DOMAIN  2536   2631       IG-LIKE C2-TYPE DOMAIN 11.
FT   DOMAIN  2632   2728       IG-LIKE C2-TYPE DOMAIN 12.
FT   DOMAIN  2729   2828       IG-LIKE C2-TYPE DOMAIN 13.
FT   DOMAIN  2829   2926       IG-LIKE C2-TYPE DOMAIN 14.
FT   DOMAIN  2927   3023       IG-LIKE C2-TYPE DOMAIN 15.
FT   DOMAIN  3024   3114       IG-LIKE C2-TYPE DOMAIN 16.
FT   DOMAIN  3115   3213       IG-LIKE C2-TYPE DOMAIN 17.
FT   DOMAIN  3214   3300       IG-LIKE C2-TYPE DOMAIN 18.
FT   DOMAIN  3301   3401       IG-LIKE C2-TYPE DOMAIN 19.
FT   DOMAIN  3402   3490       IG-LIKE C2-TYPE DOMAIN 20.
FT   DOMAIN  3491   3576       IG-LIKE C2-TYPE DOMAIN 21.
FT   DOMAIN  3577   3664       IG-LIKE C2-TYPE DOMAIN 22.
FT   DOMAIN  3665   3845       LAMININ G-LIKE 1.
FT   DOMAIN  3846   3883       EGF-LIKE 1.
FT   DOMAIN  3886   3924       EGF-LIKE 2.
FT   DOMAIN  3930   4105       LAMININ G-LIKE 2.
FT   DOMAIN  4106   4143       EGF-LIKE 3.
FT   DOMAIN  4145   4178       EGF-LIKE 4.
FT   DOMAIN  4203   4391       LAMININ G-LIKE 3.
FT   SITE      65     67       HEPARAN SULFATE (POTENTIAL).
FT   SITE      71     73       HEPARAN SULFATE (POTENTIAL).
FT   SITE      76     78       HEPARAN SULFATE (POTENTIAL).
FT   SITE    4151   4153       MEDIATES MOTOR NEURON ATTACHMENT
FT                             (POTENTIAL).
FT   SITE    4301   4303       MEDIATES MOTOR NEURON ATTACHMENT
FT                             (POTENTIAL).
FT   DISULFID  199    212      BY SIMILARITY.
FT   DISULFID  206    225      BY SIMILARITY.
FT   DISULFID  219    234      BY SIMILARITY.
FT   DISULFID  285    297      BY SIMILARITY.
FT   DISULFID  292    310      BY SIMILARITY.
FT   DISULFID  304    319      BY SIMILARITY.
FT   DISULFID  325    337      BY SIMILARITY.
FT   DISULFID  332    350      BY SIMILARITY.
```

```
Query Match             5.9%;    Score 122.5;    DB 1;    Length 4393;
Best Local Similarity   62%;     Conservative 25.3%;    Pred. No. 0.26;
Matches    62;    Conservative 28;    Mismatches 80;    Indels 75;    Gaps 12;

Qy      6  LTVPEAVGSG--SGNEVIE-GPQNATVLKGSQARFNCTVSQG------WKLIMWALSDM   55
           :: :: || | | | :| ||| :|| || |||| ||        || | |||  |
Db   3009  VTVPPSEGSSVRLRSPEVISIDPPSSTVQQGQODASPKCLIHDGAAPISLEWKTRNQELEDN 3068

Qy     56  VVLSVRPMEPLIITNDRFTSQORYDQGGNFTSEMIIHNVEPSDSGNIRC-------SLQ  105
           :: ||    |    | ||                ||:  | | ::         :|
Db   3069  VHIS--PNGSIIT-----------------------IVGTRPSNHGTYRCVASNAYGVAQSVV 3106

Qy    106  NSRLHGSAYLIVQVMGELFI---PSVNL-VVAENEPCEVTCLPSHWTRLPDISWEL----  157
           | || | :|| || |   | :|   | |:|   | | |  ||  ||
Db   3107  NLSVHGPPTVSVLPEGPVWWKVGKRAVTLECVSAGEPRS----SARWTRISSTPAKLEQRT 3162

Qy    158  -GLLVSHSSYYFVPEPSDLQSAVSILALPQSNGTLTCVATWKSLKARKSATVNLTVIRC  216
            ||  :| ||  :|  :|  || |  :|      | | |:     :| |:|||| :|  :|
Db   3163  YGLMDSH------------TVLQISSAKPSDAGTYVCLAQNALGTAQKQVEVIV-----  3204

Qy    217  PQDTG 221
           |||
Db   3205  --DTG 3207
```

RESULT 13
SMP_COTJA

```
ID   SMP_COTJA     STANDARD;      PRT;   620 AA.
AC   Q92154;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Schwann cell myelin protein precursor.
GN   SMP.
OS   Coturnix coturnix japonica (Japanese quail).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Coturnix.
OX   NCBI_TaxID=93934;
```

```
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 18-28; 120-132; 135-157 AND
RP   563-571.
RA   Dulac C., Tropak M.B., Cameron-Curry P., Rossier J., Marshak D.R.,
RA   Roder J., le Douarin N.M.;
RT   "Molecular characterization of the Schwann cell myelin protein, SMP:
RT   structural similarities within the immunoglobulin superfamily.";
RL   Neuron 8:323-334(1992).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EXCLUSIVELY EXPRESSED BY MYELINATING AND
CC       NONMYELINATING SCHWANN CELLS AND OLIGODENDROCYTES.
CC   -!- DEVELOPMENTAL STAGE: FIRST SYNTHESIZED AT EMBRYONIC DAY 5, IT
CC       REMAINS EXPRESSED BY CULTURED SCHWANN CELLS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; S83711; AAB21466.1; -.
DR   HSSP; P56276; ITLK.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00047; ig; 2.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IG2; 2.
KW   Myelin; Glycoprotein; Cell adhesion; Transmembrane; Signal; Repeat;
KW   Immunoglobulin domain.
FT   SIGNAL         1     17
FT   CHAIN         18    620       SCHWANN CELL MYELIN PROTEIN.
FT   DOMAIN        18    516       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     517    536       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       537    620       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN        28    106       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN       151    223       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN       253    311       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN       339    398       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN       424    495       IG-LIKE C2-TYPE DOMAIN 4.
FT   DISULFID      35    164       BY SIMILARITY.
FT   DISULFID      40     99       BY SIMILARITY.
FT   DISULFID     158    216       BY SIMILARITY.
FT   DISULFID     260    304       BY SIMILARITY.
FT   DISULFID     346    391       BY SIMILARITY.
FT   DISULFID     420    429       BY SIMILARITY.
FT   DISULFID     431    488       BY SIMILARITY.
FT   CARBOHYD     222    222       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     314    314       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     331    331       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     405    405       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD     449    449       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   620 AA;  66943 MW;  004B3EC7EDC18FBA CRC64;

Query Match             5.7%;   Score 118.5;  DB 1;  Length 620;
Best Local Similarity  21.5%;   Pred. No. 0.049;
Matches   57;  Conservative  33;  Mismatches  88;  Indels  87;  Gaps  12;

Qy    9 PEAVGSGSGNEVIEGPQONATVLKGSQARFNCTV-SQGWKLIMWALSDMVVLSVRPMEPII 67
         |:||        ||:|   |  |||   :|       |||
Db  239 PQVGLWGPTEVVE------GSDVELGCEAEGRPAPLISWFRGSEVL----------- 279

Qy   68 TNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPS 127
         |                     :  |
Db  280 ----------REEPGRNL--RLLSNVGPDGGSFSCVAENRHGRINRSLQLRVAYAPRAPV 329

Qy  128 VN--LVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAV---SIL 182
```

```
RESULT 14
NEO1_CHICK
ID   NEO1_CHICK     STANDARD;      PRT;   1443 AA.
AC   Q90610;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neogenin (Fragment).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Embryonic brain;
RX   MEDLINE=95105243; PubMed=7805578;
RA   Vielmetter J., Roman J.M., Dreyer W.J.;
RT   "Neogenin, an avian cell surface protein expressed during terminal
RT   neuronal differentiation, is closely related to the human tumor
RT   suppressor molecule deleted in colorectal cancer.";
RL   J. Cell Biol. 127:2009-2020(1994).
CC   -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC       TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC       DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC       MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC   -!- DEVELOPMENTAL STAGE: IN RETINA, EXPRESSED ON GANGLION CELL FIBERS
CC       AS SOON AS THEY BEGIN TO EXTEND THEIR AXONS.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; U07644; AAC59662.1; -.
DR   HSSP; P11276; 2MFN.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FNIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 6.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 2.
DR   SMART; SM00408; IG2; 2.
KW   Transmembrane; Immunoglobulin domain; Glycoprotein; Repeat.
FT   NON_TER       1
FT   DOMAIN       <1   1090
FT   TRANSMEM   1091   1111       POTENTIAL.
FT   DOMAIN     1112   1443       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       33    102       IG-LIKE C2-TYPE DOMAIN 1.

Db  330 INGSLWVVSGDPVSVTCRAE---------------------SEPAILTVLRGGKVM 365
         :|  ||   ||                              ||:| :  ::
Qy  183 ALF------------PQSNGTLTLCTVATWKSLKARKSATVNLTV------------IRCPQD 219

Db  366 AAAIYEDHVTMEMRPARPEDGGTYSCVA--ENQHGASSTSFNISVEYPPLVLPASRCTAG 423
         :  :|:::|  |    ||||  ||| |    ::|||       :|  ::   ||
Qy  220 TGGGINIFGVLSSLP--SLGFSLPT 242

Db  424 -GDGVRCVCMVNSIPDSSLVFELPT 447
```

RESULT 15

```
ID   NEO1_HUMAN     STANDARD;      PRT;  1461 AA.
AC   Q92859; O00340;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Neogenin precursor.
GN   NEO1 OR NGN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97236653; PubMed=9121761;
RA   Meyerhardt J.A., Look A.T., Bigner S.H., Fearon E.R.;
RT   "Identification and characterization of neogenin, a DCC-related
RT   gene.";
RL   Oncogene 14:1129-1136(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Fetal brain;
RX   MEDLINE=97312699; PubMed=9169140;
RA   Vielmetter J., Chen X.-N., Miskevich F., Lane R.P., Yamakawa K.,
RA   Korenberg J.R., Dreyer W.J.;
RT   "Molecular characterization of human neogenin, a DCC-related protein,
RT   and the mapping of its gene (NEO1) to chromosomal position 15q22.3-
RT   q23.";
RL   Genomics 41:414-421(1997).
CC   -!- FUNCTION: MAY BE INVOLVED AS A REGULATORY PROTEIN IN THE
CC       TRANSITION OF UNDIFFERENTIATED PROLIFERATING CELLS TO THEIR
CC       DIFFERENTIATED STATE. MAY ALSO FUNCTION AS A CELL ADHESION
CC       MOLECULE IN A BROAD SPECTRUM OF EMBRYONIC AND ADULT TISSUES.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; 1 (SHOWN HERE) AND 2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED AND ALSO IN CANCER CELL
CC       LINES.
CC   -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. DCC
CC       SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 4 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 FIBRONECTIN TYPE III-LIKE DOMAINS.
```

```
DR   EMBL; U61262; AAB17263.1; -.
DR   EMBL; U72391; AAC51287.1; -.
DR   HSSP; P02751; 1TTF.
DR   MIM; 601907; -.
DR   Genew; HGNC;7754; NEO1.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_C2.
DR   InterPro; IPR003600; Ig_like.
DR   Pfam; PF00041; fn3; 6.
DR   Pfam; PF00047; ig; 4.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 6.
DR   SMART; SM00410; IG_like; 1.
DR   SMART; SM00408; IGC2; 3.
KW   Transmembrane; Immunoglobulin domain; Glycoprotein; Signal; Repeat;
KW   Alternative splicing.
FT   SIGNAL        1     33       POTENTIAL.
FT   CHAIN        34   1461       NEOGENIN.
FT   DOMAIN       34   1105       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1106   1126       POTENTIAL.
FT   DOMAIN     1127   1461       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       67    136       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      166    228       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      263    327       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      355    417       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      436    533       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      536    629       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      630    729       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      735    829       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      850    950       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      951   1052       FIBRONECTIN TYPE-III 6.
FT   DOMAIN     1118   1121       POLY-VAL.
FT   DISULFID     74    129       BY SIMILARITY.
FT   DISULFID    173    221       BY SIMILARITY.
FT   DISULFID    270    320       BY SIMILARITY.
FT   DISULFID    362    410       BY SIMILARITY.
FT   CARBOHYD     73     73       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    210    210       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    326    326       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    470    470       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    489    489       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    639    639       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    715    715       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    909    909       N-LINKED (GLCNAC...) (POTENTIAL).
FT   VARSPLIC   1248   1300       MISSING (IN ISOFORM 2).
FT   CONFLICT    168    168       G -> N (IN REF. 2).
SQ   SEQUENCE   1461 AA;  159958 MW;  7AAE897E69635A21 CRC64;
```

```
FT   DOMAIN      132    194       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      229    293       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      321    383       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      422    519       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      522    615       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      616    714       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      720    814       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      835    935       FIBRONECTIN TYPE-III 5.
FT   DOMAIN      936   1037       FIBRONECTIN TYPE-III 6.
FT   DISULFID     40     95       BY SIMILARITY.
FT   DISULFID    139    187       BY SIMILARITY.
FT   DISULFID    236    286       BY SIMILARITY.
FT   DISULFID    328    376       BY SIMILARITY.
FT   CARBOHYD     39     39       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    176    176       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    292    292       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    456    456       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    475    475       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    625    625       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    700    700       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    894    894       N-LINKED (GLCNAC...) (POTENTIAL).
SQ   SEQUENCE   1443 AA;  158050 MW;  55BC67955790E826 CRC64;
```

```
Query Match             5.7%;  Score 117.5;  DB 1;  Length 1443;
Best Local Similarity   26.2%;
Matches   49;  Conservative   28;  Mismatches   73;  Indels   37;  Gaps   11;
                        Pred. No. 0.17;

Qy         21 IEGPQNATVLKGSQARFNCTVSQGW--KLIMVALSDMVVLSVRPMEPIITNDRFTSQRY- 77
              ::: :  ::  :: :  :  :  : :   :::  :  :    : : :::  :  :
Db        218 VROPSSLTKVTGQNAVFPC-VAGGFPTPYVRWTKNG----------EELITED---SERFA 264

Qy         78 -DGQGNFTFSEMIIHNVEPSDSGNIRC--SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAE 134
               : :  :  :       :  :  :: :   : : ::  :: : :  ::  ::  :: ::
Db        265 LRAGGS----LLISDVTEEDVGTYTCIADNENETIEAQAELAVQVPPEFLKRPANIYAHE 320

Qy        135 NE----PCEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQSN 189
               :    :     :: :  :       :::: :            : :  : : :
Db        321 SMDIVFEEVTGKPT------PTVKWVKNGDVVIPSDYFKIVKEHNLQ----VLGLVKSDE 371

Qy        190 GILTCVA 196
              : :: :
Db        372 GFYCIA 378
```

```
Query Match           5.5%;   Score 114.5;  DB 1;  Length 1461;
Best Local Similarity 24.4%;   Pred. No. 0.3;
Matches   68; Conservative  31; Mismatches  107; Indels  73; Gaps  13;

Qy   16 SGNEVIEGPQQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITN------ 69
        || ||| ||    | || ||| | :  |:::
Db  203 SGMLVI---SNAT--EGDGGLYRCVVESGGPP--KYSDEVELKVLPDPEVISDLVFLKQ 254

Qy   70 ------------------------DRFTSQRYD--QGGNFTSEMIIHNV 92
                                |:|:|  ||:   | :|
Db  255 PSPLVRVIGQDVVLPCVASGLPTPTIKWMKNEEALDTESSERLVLLAGGS-----LEISDV 310

Qy   93 EPSDSGNIRCSLQNSR--LHGSAYLTVQVWGELFIPSVNLVVAENE-----PCEVTCLPSH 146
        || || |||  | :    | :  |   ||||               ||||:|:
Db  311 TEDDAGTYFCIADNGNETIEAQAELTVQQPEFLKQPTNIYAHESMDIVFECEVTGKPT- 369

Qy  147 WTRLPDISW-ELGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARK 205
        | : | :  | : | || | :|| |   :|  | :|   |     | | || | |
Db  370 ----PTVKWVKNGDMWIPSDYFKIVKEHNLQ-----VLGLVKSDEGFYQCIAENDVGNAQA 421

Qy  206 SATVNLTVIRCPQDTGGGINIP-GVLSSLPSLGFSLPTW 243
        |:|:      |:|||||| ||   ||
Db  422 GAQL-IILEHAPATTGPLPSAPRDVVASLVSTRFIKLTW 459

Search completed: April 28, 2003, 18:09:47
Job time : 16.3692 secs
```

OM protein - protein search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.4_p5_4578

Run on: April 28, 2003, 18:09:54 ; Search time 19.0659 Seconds
(without alignments)
1946.300 Million cell updates/sec

Title: US-09-729-264-6
Perfect score: 2077
Sequence: 1 MERHLLTVPEAVGSGSGNEV..........HPQASFNLASPEKVSNTTVV 386

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : PIR_73:*
            1: pir1:*
            2: pir2:*
            3: pir3:*
            4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	151	7.3	404	1	I61596	advanced glycosyla
2	139.5	6.4	344	2	I56551	neurotrimin - rat
3	133	6.4	1427	2	I51669	tumor suppressor -
4	130	6.3	345	2	JC4025	opioid-binding cel
5	129.5	6.3	164	2	T24272	hypothetical prote
6	129	6.2	345	2	S03199	opioid-binding pro
7	129	6.2	6642	2	T29757	protein UNC-89 - C
8	128	6.2	345	2	JC1239	opioid-binding pro
9	128	6.2	1447	2	A54100	tumor suppressor p
10	127.5	6.1	871	1	I48696	protein-tyrosine k
11	127.5	6.1	881	1	I48697	protein-tyrosine k
12	125	6.0	1070	2	JC4593	protein-tyrosine k
13	122.5	5.9	188	2	T15651	hypothetical prote
14	122.5	5.9	4391	2	A38096	perlecan precursor
15	118.5	5.7	338	2	JC1238	opioid-binding pro
16	118.5	5.7	620	2	JH0593	Schwann cell myeli
17	117.5	5.7	1443	2	I50600	neogenin - chicken
18	114.5	5.5	152	2	T18975	hypothetical prote
19	114.5	5.5	416	1	A42879	advanced glycosyla
20	111	5.5	693	2	S49228	sodium-dependent p
21	111	5.3	764	2	A49448	irregular chiasm C
22	110.5	5.3	1177	2	T16594	hypothetical prote
23	109.5	5.3	890	1	A53743	protein-tyrosine k
24	109	5.2	391	2	T09058	butyrophilin homol
25	108.5	5.2	487	2	S65133	butyrophilin - mou
26	108.5	5.1	802	2	T13149	butyrophilin homol
27	106.5	5.1	946	1	A47299	mitogen-and stress
28	106.5	5.1	2491	1	A28372	ror-related recept
29	106	5.1	1092	1	JN0635	neural cell adhesi
30	105.5	5.1	862	2	I49583	differentiation an
31	105.5	5.1	882	2	I38912	receptor tyrosine
32	105.5	5.1	3707	2	S18252	heparan sulfate pr
33	105	5.1	662	2	T16525	hypothetical prote
34	105	5.1	5825	2	T12117	polyprotein - fava
35	104	5.0	2222	2	T13924	sdk protein - frui
36	103.5	5.0	423	2	T29549	hypothetical prote
37	103.5	5.0	1323	2	PN0568	connectin 3B - chi
38	103	5.0	365	2	JC7780	CD22 homolog/B lym
39	103	5.0	1273	2	T42405	T-cell surface gly
40	102	4.9	721	2	T41530	coxsackie- and ade
41	102	4.9	868	2	A46512	sax-3 protein - Ca
42	101.5	4.9	344	1	RWRTC2	hypothetical prote
43	101.5	4.9	421	2	T46266	hypothetical prote
44	101.5	4.9	841	2	JC5894	killer cell inhibi
45	101	4.9	338	2	JC5519	50K glycoprotein p

ALIGNMENTS

RESULT 1
I61596
advanced glycosylation end-products receptor precursor - human
N;Alternate names: advanced glycosylation end-products receptor; product-binding protein, 35K; glycoprot
C;Species: Homo sapiens (man)
C;Date: 24-May-1996 #sequence_revision 07-Feb-1997 #text_change 16-Jul-1999
C;Accession: I61596; B42879; S27968
R;Sugaya, K.; Fukagawa, T.; Matsumoto, K.; Mita, K.; Takahashi, E.; Ando, A.; Inoko,
Genomics 23, 408-419, 1994
A;Title: Three genes in the human MHC class III region near the junction with the cla
nterpart of mouse mammary tumor gene int-3.
R;Neeper, M.; Schmidt, A.M.; Brett, J.; Yan, S.D.; Wang, F.; Pan, Y.C.; Elliston, K.;
J. Biol. Chem. 267, 14998-15004, 1992
A;Title: Cloning and expression of a cell surface receptor for advanced glycosylation
A;Reference number: A42879; MUID:92340547; PMID:1378843
A;Accession: B42879
A;Molecule type: mRNA
A;Residues: 'G',2-99,'R',101-404 <NEE>
A;Cross-references: EMBL:M91211; NID:g190845; PIDN:AAA03574.1; PID:g190846
A;Experimental source: lung
A;Note: sequence extracted from NCBI backbone (NCBIP:109438)
C;Comment: Advanced glycosylation end products are heterogeneous nonenzymatically gly
cellular function, thus contributing to tissue lesions in diabetes.
C;Comment: This receptor appears also to mediate the effects of amyloid beta peptide
ates in the neurotoxic pathway that produces dementia in Alzheimer's disease.
C;Genetics:
A;Gene: GDB:AGER
A;Cross-references: GDB:306354; OMIM:600214
A;Map position: 6p21.3-6p21.3
A;Introns: 18/1; 53/3; 119/1; 140/3; 170/1; 231/1; 274/3; 322/1; 331/1; 373/2
C;Function:
A;Description: neuronal receptor for amphoterin, a DNA-binding protein involved in ne
C;Superfamily: advanced glycosylation end products receptor; immunoglobulin homology
C;Keywords: Alzheimer's disease; glycoprotein; receptor; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-404/Product: advanced glycosylation end products receptor #status predicted <MAT
F;23-344/Domain: extracellular #status predicted <EXT>
F;31-101/Domain: immunoglobulin homology <IM1>
F;137-210/Domain: immunoglobulin homology <IM2>
F;252-303/Domain: immunoglobulin homology <IM3>
F;345-362/Domain: transmembrane #status predicted <TMM>
F;363-404/Domain: intracellular #status predicted <INT>
F;25,81/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;38-99,144-208,259-301/Disulfide bonds: #status predicted

A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/
A;Molecule type: DNA
A;Reference number: I61596 <RES>

Query Match   7.3%;   Score 151;   DB 1;   Length 404;

Best Local Similarity 23.3%; Pred. No. 0.001;
Matches 90; Conservative 38; Mismatches 112; Indels 146; Gaps 19;

```
Qy      9 PEAVGSGSGNEVIEGPQNATVLKGSQARFNCTVSQG---WKLIMWALSDMVVLSVRPMEP 65
            ||||  || || |||:|: |:|:|  | |:|:       ||:| |||||||||::
Db    124 PEIVDSAS--ELTAGVPN-----KVGTCVSEGSYPAGTLSWHLDG------KP 163

Qy     66 IITNDRETS-----QRYDQGGNFTI--SEMIIHNVEPSDSGNIR----CSLQNSRLHGSAY 114
            :||::|:||        |:|| ||| ::: ::||:|||||||||    | :|::|| |||
Db    164 LVPNEKGVSVKEQTRRHPETGLFTLQSSELM--VTPARGDGRPTFSCSFSPGLPRHRAL 220

Qy    115 LTVQVMGELFIP--------SVNLVVAENEP--------CEVTCLPSHWTRLPDISWE 156
            ||||| :|||:|        |:||::|||:|        |||| ||:|||| ||:|
Db    221 RTFAPIQPRVWEPVPLEEVQLVV---EPEGGAVAPGGTVTLTCEVPAQPS-----PQIHWM 272

Qy    157 LGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRC 216
            | ||:|||:|||||||:||||:||||||||||||||||||||:|:||||||||||||:
Db    273 KD---------GVPLPLPPSPVLILPEIGPQDGGTVSCVATHSSHGPQESRAVSISIIE- 322

Qy    217 PQDTGGGTNIPGVLSSLPSLGPSLPTWGKVGLGLAGTMLLT----PTCTLTIRCCCC 269
            |  | | ||||:| || |:||  ||:|| :: |: || |::     | |:| |||||
Db    323 PGEEG---------------------PTPAGSVGGSGLGTLALALGILGGLGTRAALLGIVILW 363

Qy    270 RRRCCGCQNCCCRCCFCCRRKRGFRIQFQKKSEKEKI--NKETETESGNENSGYNSDEQKT 327
            ||: || |  | | |||:| : ||:|||:|||:||:|  :||:||||||:||:|||:|::|
Db    364 QRR-----------------ORRGEERKAPENQEEEEERAELN------------------- 389

Qy    328 TETASLPPKSCESSSDPEQRNSSCGPP 353
            ::: ||||:||| ||||||::|:||||
Db    390 ---------QSEEPEAGESSTGGP 404
```

RESULT 2
I5551
neurotrimin - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 19-May-2000
C;Accession: I5551
R;Struyk, A.F.; Canoll, P.D.; Wolfgang, M.J.; Rosen, C.L.; D'Eustachio, P.; Salzer, J.L.
J. Neurosci. 15, 2141-2156, 1995
A;Title: Cloning of neurotrimin defines a new subfamily of differentially expressed neur
A;Reference number: I55551; MUID:95198094; PMID:7891157
A;Accession: I5551
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-344 <RES>
A;Cross-references: EMBL:U16845; NID:g755184; PIDN:AAA67445.1; PID:g755185
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin

Query Match   6.7%; Score 139.5; DB 2; Length 344;
Best Local Similarity 24.9%; Pred. No. 0.0062;
Matches 66; Conservative 40; Mismatches 110; Indels 49; Gaps 13;

```
Qy      3 RHLLTVPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLS-- 59
            | |||||:| |||||||| |||||||||||||:|||||| |:|| ||||||||||||
Db     20 RLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDNRVTRVAWLNRSTILYAGN 79

Qy     60 -----VRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSRL 109
                 |||:|:||||||||||||||||||||||:|||||||||||||| :|      |||
Db     80 DKWCLDPRVVLLSN----SIEIQNVDVIDEGPYTCSVQTDNHPKTSRV 128

Qy    110 HGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYFV 169
            |||||||||||||||||||||||||||| |   |:|:|| |:  ||:| :|||||:|
Db    129 H---LIVQVSPKIVEISSDISINEGNSISITCIA?GRPE-PTVTWR--HISPRAVGFV 180

Qy    170 PEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP----QDTGGGI 224
            | |||  | || :|:|||:| | |:|  |  |:| :  | |||| ||      |||
Db    181 SEDEYLE---TQGITREQSGEYECSAS-NDVAAPVVRRVNVTVNVPPYISEAKGTGVPV 235

Qy    225 NIPGVL----SSLPSLGFSLPTWGK 245
            |:||:|    :|:|:|| ||| ||
Db    236 GQKGTLQCEASAVPSAEFQ---WFK 257
```

RESULT 3
I51669
tumor suppressor - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I51669
R;Pierceall, W.E.; Reale, M.A.; Candia, A.F.; Wright, C.V.; Cho, K.R.; Fearon, E.R.
Dev. Biol. 166, 654-665, 1994
A;Title: Expression of a homologue of the deleted in colorectal cancer (DCC) gene in
A;Reference number: I51669; MUID:95113183; PMID:7813784
A;Accession: I51669
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1427 <PIE>
A;Cross-references: EMBL:U10986; NID:g606873; PIDN:AAA70168.1; PID:g606874
C;Genetics:
A;Gene: XDCCa

Query Match   6.4%; Score 133; DB 2; Length 1427;
Best Local Similarity 25.7%; Pred. No. 0.087;
Matches 59; Conservative 33; Mismatches 90; Indels 48; Gaps 13;

```
Qy     21 IEGPQNATVLKGSQARFNCTVSQGW----KLIMWALSDMVVLSVRPMEPIITNDRFTSQRYD 78
            |||||||||||||:||||| |:||     ||||||||||||:|||:|||||||||||||||
Db    347 DIEFECAVSGKPS---------PIVKWIKNGEVVIPSDVFQIVDGSNLR---ILGLVKSDEGY 397

Qy     79 QGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAENE 136
            |||||||:|||||||||||||| |  |:|::||||||||||| ||||||||||||::||
Db    289 VLGG--SNLLISNVIDDDAGAYTCVATYKNENTSFSADLTVMVPPQFLNHPANLYAYESM 346

Qy    137 PCEVTCL----PSHWTRLPDISW-ELGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGT 191
            ||||||     |:|:|| ||:|  |:|| :|||:|||||||:||||:||||||||:|
Db    243 LQRPSNVVAIEGQDAVLECAVS-GYPTPTIVWMQGD-----EPVPIRTR----KYS 288
```

RESULT 4
JC4025
opioid-binding cell adhesion protein - human
C;Species: Homo sapiens (man)
C;Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 17-Mar-2000
C;Accession: JC4025
R;Shark, K.B.; Lee, N.M.
Gene 155, 213-317, 1995
A;Title: Cloning, sequencing and localization to chromosome 11 of a cDNA encoding a h
A;Reference number: JC4025; MUID:95337612; PMID:7721093
A;Accession: JC4025
A;Molecule type: mRNA
A;Residues: 1-345 <SHA>
A;Cross-references: GB:L34774; NID:g514373; PIDN:AAA36387.1; PID:g514374
A;Experimental source: brain
C;Comment: This protein binds opioid alkaloids in the presence of acidic lipids, exhi
C;Genetics:
A;Gene: GDB:OPCML; OBCAM; OPCM
A;Cross-references: GDB:251677; OMIM:600632
A;Map position: 11pter-11qter
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-ter

Query Match   6.3%; Score 130; DB 2; Length 345;
Best Local Similarity 25.6%; Pred. No. 0.031;
Matches 73; Conservative 34; Mismatches 108; Indels 70; Gaps 17;

```
Qy      3 RHLLTVPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVL--- 58
            | |||||:| |||||||| |||||||||||||:|||||| |:|| ||||||||||||
Db     20 RLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW--LNRSTILYAG 78

Qy     59 -----SVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSR 108
                 |||:|:|||||||||||||||||||||||:||||||||||| | |     |||
Db     79 NDKWSIDPRVILVN--TPTQY----------SIMIQNVDVYDEGPYTCSVQTDNHPKTSR 127
```

```
Qy  109 LHGSAYLITVQVMGELFIPSVNLVVAENEPCEVTCL-----PSHWTRLPDISWELGLLVSHS 164
            ::|  |:  :|   ||: : || |
Db  128 VH------LIVQVPPQIMNISSDITVNEGSSVTLCLAIGRPE-----PTVTWR-----HL 172

Qy  165 SYY-----FVPEPSDLQOSAVSILALTPQSNGTLTCVAATWKSLKARKSATVNLIIV----IR 215
            |:   ||: : | | |||
Db  173 SVKEGQGFVSEDEYLE-----ISDIKRDQSGEYECSAL-NDVAAPDVRKVKIITVNYPPYIS 227

Qy  216 CPQDTGGGINIPGVL---------SSLPSLGFSLPTWGK-----VGLGLAG 252
            :||                |:|  |   |
Db  228 KAKNTGVSVGQKGILSCEASAVPMAEFQ---WFKEETRLATGLLDG 269
```

RESULT 5
T24272
hypothetical protein T01B7.8 - Caenorhabditis elegans
R;Sims, M.
submitted to the EMBL Data Library, October 1995
A;Reference number: Z18867
A;Accession: T24272
C;Species: Caenorhabditis elegans
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-164 <WIL>
A;Cross-references: EMBL:Z66499; PIDN:CAA91301.1; GSPDB:GN00020; CESP:T01B7.8
A;Experimental source: clone T01B7
C;Genetics:
A;Gene: CESP:T01B7.8
A;Map position: 2
A;Introns: 20/3; 90/2

```
Query Match         32.1%;  Score 129.5;  DB 2;  Length 164;
Best Local Similarity 32.1%;  Pred. No. 0.015;
Matches  43;  Conservative 11;  Mismatches 53;  Indels 27;  Gaps 6;

Qy  157 LGLLVSHSSYYFVPEPSDLQOSAVSILALTPQSNGTLTCVAATWKSLKARKSATVNLTVIRC 216
            |:  ||:||:  |   ||:  :  | | |||
Db    6 LAILLAIGFIAV--SQVQSAV-----LPVSSTELATVGTDVSTASTAIDTLGNSSRV 57

Qy  217 PQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLLAGTMLLTPTCTLTIRCCCRRRCCGC 276
Db   58 KRQGGCGGCCGCGC-------------GCCGCGGGGGG--CGCCCCRPRCCCCCRRCCTC 101

Qy  277 --NCCC--RCCFCCR 287
          :|||   :|  :|
Db  102 CRTCCCTRCCTCCR 115
```

RESULT 6
S03199
opioid-binding protein OPCAM precursor - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 17-Mar-2000
C;Accession: S03199
R;Schofield, P.R.; McFarland, K.C.; Hayflick, J.S.; Wilcox, J.N.; Cho, T.M.; Roy, S.; Le...
EMBO J. 8, 489-495, 1989
A;Title: Molecular characterization of a new immunoglobulin superfamily protein with pot...
A;Reference number: S03199; MUID:89251576; PMID:2721489
A;Accession: S03199
A;Molecule type: mRNA
A;Residues: 1-345 <SCH>
A;Cross-references: EMBL:X12672; NID:g585; PIDN:CAA31192.1; PID:g586
A;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin...
C;Keywords: transmembrane protein
F;1-27/Domain: signal sequence #status predicted <SIG>
F;28-345/Product: opioid-binding protein OPCAM #status predicted <MAT>

Query Match   25.7%; Score 129; DB 2; Length 345;
Best Local Similarity 25.7%; Pred. No. 0.037;
Matches 72; Conservative 33; Mismatches 109; Indels 66; Gaps 16;
```

```
Qy    3 RHLLIVPEAVGSGSGGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVL---58
            | |:|| :| |   |:: |  ||:|||| ||  |  ||||
Db   20 RLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAG 78

Qy   59 -----SVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGSGNIRCSLQ----- 108
Db   79 NDKWSIDPRVIILVN---TPTQY-----SIMIQNVDVYDEGPYTCSVQTDNHPKTSR 127

Qy  109 LHGSAYLITVQVMGELFIPSVNLVVAENEPCEVTCL-----PSHWTRLPDISWELGLLVSHS 164
Db  128 VH------LIVQVPPQIMNISSDITVNEGSSVTLCLAIGRPE-----PTVTWR-----HL 172

Qy  165 SYY-----FVPEPSDLQSAVSILALTPQSNGTLTCVAATWKSLKARKSATVNLTIV----IR 215
Db  173 SVKEGQGFVSEDEYLE-----ISDIKRDQSGEYECSAL-NDVAAPDVRKVKIITVNYPPYIS 227

Qy  216 CPQDTGGGINIPGVL---------SSLPSLGFSLPTWGKVGLGLA 251
Db  228 KAKNTGVSVGQKGILSCEASAVPMAEFQ---WFKEDTRLA 264
```

RESULT 7
T29757
protein UNC-89 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C;Accession: T29757
R;Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid C09D1.
A;Reference number: Z20679
A;Accession: T29757
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6642 <DUZ>
A;Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
A;Experimental source: strain Bristol N2; clone C09D1
C;Genetics:
A;Gene: CESP:unc-89
A;Map position: 1
A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

```
Query Match         23.0%;  Score 129;  DB 2;  Length 6642;
Best Local Similarity 23.0%;  Pred. No. 0.89;
Matches  59;  Conservative 41;  Mismatches 97;  Indels 60;  Gaps 10;

Qy    6 LTVPEAVGSG-----SGNE--------------VIEGPQNATVLKGSQARFNCT 40
Db 2038 LIIPNAQDSGKITVEASNEVGSSESSAQLTVNPPSTTPIVVDGPKSVTIKETETAEFKAT 2097

Qy   41 VSQGW--KLIMWALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSG 98
Db 2098 IS-GFPAPTVKWIINEKIVEESRTITIIKTEDVYT-----LKISNAKIEQTG 2143

Qy   99 NIRCSLQNSRLHGSAYLITVQVMGELFIPS-----VNLVVAENEPCEVTCL--PSHWTRL 150
Db 2144 TVKVTAQNSAGQDSKQADLKVEPNVKAPKFKSQLTDKVADEGEPLRWNLELDGPSPGT-- 2201

Qy  151 PDISWEL-GLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVAATWKSLKARKSATV 209
Db 2202 -EVSMLLNGQPLTKSDTVQVVDHGDGTYHVTIAEAKPEMSGTITAKAKNAAGECETSAKV 2260

Qy  210 NLIT------VIRCPQD 219
Db 2261 TVNGGNKKPEEVQAPQN 2277
```

RESULT 8
JC1239
opioid-binding protein (clones SG8 and SG13) - rat
C;Species: Rattus norvegicus (Norway rat)

C;Date: 03-May-1994
C;Accession: JC1239
R;Lippman, D.A.; Lee, N.M.; Loh, H.H.
Gene 117, 249-254, 1992
A;Title: Opioid-binding cell adhesion molecule (OBCAM)-related clones from a rat brain c
A;Reference number: JC1238; MUID:92347701; PMID:1339369
A;Accession: JC1239
A;Molecule type: mRNA
A;Residues: 1-345 <LIP>
C;Cross-references: GB:M88710; NID:g203247; PIDN:AAA40859.1; PID:g203248; GB:M88711; NID
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin
C;Keywords: transmembrane protein

Query Match   6.2%;   Score 128; DB 2; Length 345;
Best Local Similarity   25.3%;   Pred. No. 0.044;
Matches   73; Conservative   35; Mismatches 110; Indels   70; Gaps   17;

Qy   3 RHLLTYPEAVGSGSGNEVI-EGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVL--- 58
          :|: |||| :: |||: |||||||||||||||||||||||||||:|||||||||
Db   20 RLLFLVPTGVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW-LNRSTLIYAG 78

Qy   59 ---SVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ------NSR 108
            |||||||||||||||||||||||||||||||||||||||||||| |||
Db   79 NDKWSIDPRVTLLVN---TPTQY---------------SIMIQNVDVDEGPYTCSVQTDNHPKTSR 127

Qy 109 LHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCI------PSHWTRLPDISWELGLLVSHS 164
         ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db 128 VH----LIVQVPPQIMNISSDITVNEISSVTLLCLAIGRPE----PTVTWR------HL 172

Qy 165 SYY---FYPEPSDLQSAVSILALTPQSNGTILTCVATWKSLKARKSATVNLTV----- 215
          ||| |||
Db 173 SVKEQGFVSEDEYLE---ISDIKRDQSGEYECSAL-NDVAAPDVRKITVNYPPYIS 227

Qy 216 CPQDTGGGINIPGVL-----SSLPSIGFSLPTWGK----VGLGLAGTML 255
         || |
Db 228 KAKNTGVSVGQKILSCEASAVPMAEFQ---WFKEDTRLATGLDGVRI 272

RESULT 9
A54100
tumor suppressor protein DCC precursor - human
N;Alternate names: colorectal cancer suppressor DCC
C;Species: Homo sapiens (man)
C;Date: 02-Aug-1994 #sequence_revision 02-Aug-1994 #text_change 05-Nov-1999
C;Accession: A54100; A40098
R;Hedrick, L.; Cho, K.R.; Fearon, E.R.; Wu, T.C.; Kinzler, K.W.; Vogelstein, B.
Genes Dev. 8, 1174-1183, 1994
A;Title: Identification of a chromosome 18q gene that is altered in colorectal cancers.
A;Reference number: A40098; MUID:90100559; PMID:2294591
A;Accession: A40098
A;Molecule type: mRNA
A;Residues: 1-750 <FEA>
C;Cross-references: GB:M32292; NID:g181492; PIDN:AAA35751.1; PID:g181493
C;Genetics:
A;Gene: GDB:DCC
A;Cross-references: GDB:119838; OMIM:120470
A;Map position: 18q21.1-18q21.1
C;Keywords: transmembrane protein; tumor suppressor
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-1447/Product: tumor suppressor protein DCC #status predicted <MAT>

Query Match   6.2%;   Score 128; DB 2; Length 1447;
Best Local Similarity   24.8%;   Pred. No. 0.21;
Matches   60; Conservative   33; Mismatches   99; Indels   50; Gaps   10;

A;Title: The DCC gene product in cellular differentiation and colorectal tumorigenesis.
R;Fearon, E.R.; Cho, K.R.; Nigro, J.M.; Kern, S.E.; Simons, J.W.; Ruppert, J.M.; Hamilto
Science 247, 49-56, 1990

RESULT 10
I48696
protein-tyrosine kinase (EC 2.7.1.112) nsk2 precursor, splice form 4 - mouse
N;Alternate names: receptor-type tyrosine kinase
N;Contains: protein-tyrosine kinase nsk2 precursor, splice form 2 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C;Accession: I48696; S60738
R;Ganju, P.; Walls, E.; Brennan, J.; Reith, A.D.
Oncogene 11, 281-290, 1995
A;Title: Cloning and developmental expression of Nsk2, a novel receptor tyrosine kina
A;Reference number: I48696; MUID:95349951; PMID:7624144
A;Accession: I48696
A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-871 <GAN1>
A;Cross-references: MGI:103308
A;Gene: nsk2
A;Genetics:
C;Comment: For alternate splice forms see PIR:I48697.
C;Experimental source: splice form 4
A;Cross-references: EMBL:X86444; NID:g929723
A;Experimental source: splice form 2
A;Cross-references: EMBL:X86444; NID:g929723
A;Molecule type: DNA
A;Residues: 1-456; 'A', 466-871 <GAN2>
A;Cross-references: EMBL:X86444; NID:g929723; PIDN:CAA60165.1; PID:g929724
A;Superfamily: mouse ror-related receptor; immunoglobulin homology; protein kinase ho
C;Keywords: ATP; glycoprotein; phosphotransferase; receptor; transmembrane protein; t
F;1-21/Domain: signal sequence #status predicted <SIG>
F;22-871/Product: protein-tyrosine kinase nsk2, splice form 2 #status predicted <MAT2
F;22-456/'A',466-871/Product: protein-tyrosine kinase nsk2, splice form 4 #status pre
F;42-101/Domain: immunoglobulin homology <IMM1>
F;135-192/Domain: immunoglobulin homology <IMM2>
F;226-284/Domain: immunoglobulin homology <IMM3>
F;498-518/Domain: transmembrane #status predicted <TRM>
F;575-865/Domain: protein kinase homology <KIN>
F;583-591/Region: protein kinase ATP-binding motif
F;222,462/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   6.1%;   Score 127.5; DB 1; Length 871;
Best Local Similarity   21.7%;   Pred. No. 0.13;
Matches   70; Conservative   41; Mismatches 125; Indels   87; Gaps   16;

Qy   14 SGSGNEV------------------------IEGPQNATVLKGSQARFNCTVSQGW--- 55
         | |||| : ||||||||||||||||:||||||
Db 220 SRTGNEAEVRILSDPGLHRQLYFLQRPSNVVAIEGKDAVLECCVS-GYPPPSFTWLRGEE 278

Qy   56 VVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSA 113
         |: ||||||||||||||||||||| ||:||||||
Db 279 VI-----------QLRSKKYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNENISASA 323

Qy 114 YLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISW-ELGLLVSHSSYFVPEP 172
         ||||||||||||||||||||||||| ::: || ||: ||:|||||| |:
Db 324 ELTVLVPPWFLNHPSNLYAYESMDIEFECTVS-GKPVPTVNWMKNGDVVIPSDYFQIVGG 382

Qy 173 SDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLIVIRCPQDTGGGINIPGVLSS 232
         || | ||: || || | | :||| ||| ||| || ||||||||||
Db 383 SNLR-----ILGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS-------VLPS 430

Qy 233 LP 234
         ||
Db 431 AP 432

Db 172 ----------SGSLRIHNVQKEDAQQYRCVAKNSL--GTAYSKLVKLEVEVLGRILRAPESHNV 223
Qy   78 DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY------LTVQVMGELFIPSVNLVV 132
           ||||||||||||||||||||||||| || ||:| |:|||||||||||
Db 122 KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD---------NALRENSRIAALE- 171
Qy   19 EVIEGPQNATVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY 77
           :||||||||||||||||||||||||| |||||||||||||||||||||||||||||:
Db

Query Match   6.1%;   Score 127.5; DB 1; Length 871;
Best Local Similarity   21.7%;   Pred. No. 0.13;
Matches   70; Conservative   41; Mismatches 125; Indels   87; Gaps   16;

```
RESULT 11
I48697
protein-tyrosine kinase (EC 2.7.1.112) nsk2 precursor, splice form 1 - mouse
N;Alternate names: receptor-type tyrosine kinase nsk2 precursor, splice form 3
C;Contains: protein-tyrosine kinase nsk2 precursor, splice form 3
C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C;Species: Mus musculus (house mouse)
A;Accession: I48697; S60740
R;Ganju, P.; Walls, E.; Brennan, J.; Reith, A.D.
A;Title: Cloning and developmental expression of Nsk2, a novel receptor tyrosine kinase
A;Reference number: I48696; MUID:95349951; PMID:7624144
Oncogene 11, 281-290, 1995
A;Accession: I48697
A;Status: nucleic acid sequence not shown; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-881 <GAN1>
A;Cross-references: EMBL:X86445; NID:g929725; PIDN:CAA60166.1; PID:g929726
A;Experimental source: splice form 1
A;Accession: S60740
A;Molecule type: DNA
A;Residues: 1-456, 'A', 466-881 <GAN2>
A;Cross-references: EMBL:X86445; NID:g929725
A;Experimental source: splice form 3
A;Comment: For alternate splice forms see PIR:I48696.
A;Genetics:
A;Gene: nsk2
A;Cross-references: MGI:103308
C;Superfamily: mouse ror-related receptor; immunoglobulin homology; protein kinase homol
C;Keywords: ATP; glycoprotein; phosphotransferase; receptor; transmembrane protein; tyr
F;1-2/Domain: signal sequence #status predicted <SIG>
F;22-881/Product: protein-tyrosine kinase nsk2, splice form 1 #status predicted <MAT>
F;22-456, 'A', 466-881/Product: protein-tyrosine kinase nsk2, splice form 3 #status predi
F;42-101/Domain: immunoglobulin homology <IMM1>
F;135-192/Domain: immunoglobulin homology <IMM2>
F;226-284/Domain: immunoglobulin homology <IMM3>
F;498-518/Domain: transmembrane #status predicted <TRM>
F;575-865/Domain: protein kinase homology <KIN>
F;583-591/Region: protein kinase ATP-binding motif
F;222,462/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             6.1%;  Score 127.5;  DB 1;  Length 881;
  Best Local Similarity   21.8%;  Pred. No. 0.13;
  Matches   70;  Conservative  41;  Mismatches 125;  Indels  87;  Gaps 16;

Qy     19 EVIEGPQNATVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY 77
              ||| || |:   | |    |:||      ||| ::| |   |  :| ||   : | |
Db    122 KITRPINVKIEGLKAVLPCTTMGNPKPSVSWIKGD------NALRENSRIAALE- 171

Qy     78 DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY----LTVQVMGELFIPSVNLVV 132
              | ||||  :| | : | || |  |:| |  |::       || | |  |  |:||
Db    172 -----SGSLRIHNVQKEDAQYRCVAKNSL-GTAYSKLVKLEVVLGRILRAPESHNV 223

Qy    133 AENEPCEVTC---LPSHWTRLPDISW-ELGLLVSHSSSYFVPEPSDLQSAVSIIALTPQ 187
              ||| ||  |   || |   | |:|  |  ||:|  |: ||||||||| ||||  :|
Db    224 TFGSFVTLRCTEIGIP-----VPTISWIENGNAVSSGSGIQESVKDRVIDSRLQLFITKP- 277
```

```
RESULT 12
JC4593
protein-tyrosine kinase-related receptor protein tyrosine kinase-like protein PTK7 precursor - human
N;Alternate names: receptor protein tyrosine kinase-like protein (RPTK)
C;Species: Homo sapiens (man)
C;Date: 16-Apr-1996 #sequence_revision 24-May-1996 #text_change 24-Sep-1999
C;Accession: JC4593
R;Park, S.K.; Lee, H.S.; Lee, S.T.
J. Biochem. 119, 235-239, 1996
A;Title: Characterization of the human full-length PTK7 cDNA encoding a receptor prot
A;Reference number: JC4593; MUID:97037064; PMID:8882711
A;Accession: JC4593
A;Molecule type: mRNA
A;Residues: 1-1070 <PAR>
A;Cross-references: GB:U40271; NID:g1322231; PIDN:AAC50484.1; PID:g1322232
A;Comment: This protein is a member of receptor protein tyrosine kinase family, but p
A;Genetics:
A;Gene: GDB:PTK7
A;Cross-references: GDB:134760; OMIM:601890
A;Map position: 6p21.1-6p12.2
C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C;Keywords: cell adhesion; extracellular protein; glycoprotein; phosphotransferase; t
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1070/Product: protein-tyrosine kinase 7 #status predicted <MAT>
F;31-703/Domain: extracellular #status predicted <EXC>
F;704-725/Domain: transmembrane #status predicted <TMM>
F;726-1070/Domain: intracellular #status predicted <INC>
F;794-1065/Domain: protein kinase homology <KIN>
F;116,175,184,214,268,283,405,463,567,646/Binding site: carbohydrate (Asn) (covalent)

  Query Match             6.0%;  Score 125;  DB 2;  Length 1070;
  Best Local Similarity   24.8%;  Pred. No. 0.25;
  Matches   59;  Conservative  34;  Mismatches 109;  Indels  36;  Gaps 10;

Qy     11 AVGSGSGNEVIEGPQNATVLKGSQARFNCTVS-QGWKLIMWALSDMVVLSVRPMEPIITN 69
              ::       |  :  | |        |         ||| ::| |   |  :| ||
Db    218 SIADSSFARVVIAPQDVVVARYEAMFHCQFSAQPPSLQWLFEDETPITNRSRPPHLRR 277

Qy     70 DRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSR-----LHGSAYLTVQVMGELF 124
               |:|| |   | ||   :| | : | || |  |:| |      |  ||| || | |
Db    278 ATVFA-------NGSLLITQVRPRNAGIYRCIGQGQGRGPPIILEATLHLAEIEDMPLF 328

Qy    125 IPSVNLVVAENEPCEVTCLPSHWTRLPDISWE-LGL-LVSHSSSYFVPEPSDLQSAVSTI 182
              | |:|   ||| ||  | || |   | |:|  :|| |||| |: ||||||||| ||||
Db    329 EPRVFTAGSEE---RVTCLEPKGLPEPSVWWEHAGVRLPTHGRVI------QKGHELVL 378

Qy    183 ALTPQSN-GTLTCVATWKSLKARKSATVNLTV------IRCPQDTGGGINIPGVLSSL 233
               |  | :  ||||| |   |:|| || ||| |      | || | |||||| | || |
Db    379 ANIAESDAGYVTCHAA--NLAGQRQDVNITVA-TVPSWLKKPQDSQLEEGKPGYLDCL 434
```

```
RESULT 13
T15651
hypothetical protein C27A2.5 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15651
R;Nhan, M.
submitted to the EMBL Data Library May 1996
A;Description: The sequence of C. elegans cosmid C27A2.
```

Qy   133 AENEPCEVTC---LPSHWTRLPDISW-ELGLLVSHSSSYFVPEPSDLQSAVSIIALTPQ 187
Db   224 TFGSFVTLRCTEIGIP-----VPTISWIENGNAVSSGSGIQESVKDRVIDSRLQLFITKP- 277

Qy   188 SNGTLTCVAT---WKSLKARKSATVNLIVIR---------CPQDTGGGINIPG-----V 229
Db   278 --GLYTCIATNKHGEKFSTAKAAATVSIAEWSKSQKDSQGYCAQYRGEGVLMQGPGEKML 335

Qy   230 LSSLPSLGFSLP------------TWGKVGLGLAGTMLLTPPCTLITRCCCCRRCCGCN-- 277
Db   336 LVFLPTTSHRDPEDAQELLIHTAWNEL------KAVSPLCRPAAEALLCYHLFLECSPG 388

Qy   278 ------------CCCRCC------FCCR 287
Db   389 VVPTMPICREYCLAVKELFCAK 411

RESULT 14

A38096

perlecan precursor - human

N;Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate prote
C;Species: Homo sapiens (man)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 05-Nov-1999
C;Accession: A38096; S19256; S77946; A41059; A40306; B33625; A33625; A41736
R;Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J;Biol. Chem. 267, 8544-8557, 1992
A;Title: Primary structure of the human heparan sulfate proteoglycan from basement membr
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A;Reference number: A41736; MUID:92112994; PMID:1730768
A;Accession: S19256
A;Molecule type: mRNA
A;Residues: 1-4391 <MUR>
A;Cross-references: GB:M85289; NID:g184426; PIDN:AAA52700.1; PID:g184427
R;Kallunki, P.; Tryggvason, K.
J;Cell Biol. 116, 559-571, 1992
A;Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD prot
ell adhesion molecules, and epidermal growth factor.
A;Reference number: A41736; MUID:92112994; PMID:1730768
A;Accession: S19256
A;Molecule type: mRNA
A;Residues: 1-4391 <MUR>
A;Reference number: A38096; MUID:92235084; PMID:1569102
A;Molecule type: mRNA
A;Residues: 1-57,'D','59-434,'A','436,'FL','438-449,'Q','451-502,'A','503-792,'K','794-908,'R',
71-2979,'H','2981-2994,'G','2996-3167,'T',3169-3240,'R',3242-3426,'Q',3428-3631,'Q',3633-4
A;Cross-references: EMBL:X62515
R;Tryggvason, K.
submitted to the EMBL Data Library, October 1991
A;Reference number: S77946
A;Accession: S77946
A;Molecule type: mRNA
A;Residues: 1-57,'D','59-434,'A','436,'FL','438-449,'Q','451-502,'A','503-792,'K','794-908,'R',
71-2979,'H','2981-2994,'G','2996-3167,'T',3169-3240,'R',3242-3426,'Q',3428-3631,'Q',3633-4
A;Cross-references: EMBL:X62515
R;Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A;Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of the g
A;Reference number: A41059; MUID:92206060; PMID:1685141
A;Accession: A41059
A;Molecule type: mRNA
A;Residues: 'RT',892-908,'R',910-1101,'L',1103-1132,'L',1134-1221,'L',1223-1397 <KA2>
A;Cross-references: GB:S76436; NID:g243370; PIDN:AAB21121.1; PID:g243371
R;Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo,
Genomics 10, 673-680, 1991
A;Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cellula
A;Accession: A40306; MUID:91365376; PMID:1679749
A;Residues: 1018-1405,'G',1407-1409,'G',1411-1465 <DOD>
A;Cross-references: GB:M64283; NID:g184424; PIDN:AAA52699.1; PID:g184425
R;Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van den
J;Cell Biol. 109, 3199-3211, 1989

A;Reference number: 218382
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Accession: T15651
A;Molecule type: DNA
A;Residues: 1-188 <NHA>
A;Cross-references: EMBL:U58760; NID:g1330384; PID:g1330389; PIDN:AAB00710.1; GSPDB:GN0C
A;Experimental source: strain Bristol N2; clone C27A2
A;Molecule type: protein
A;Residues: 2166-2171,'X',2173-2175,'X',2177-2185 <HE3>
A;Gene: CESP:C27A2.5
A;Genetics:
A;Map position: 2
A;Introns: 19/3; 91/2

Qy   260 CTLTIRCCCCRRCCGG--NCCC-RCCFCCR 287
          ||||| |||| |||| ||| ||| |||
Db   86 CCRPKCCCCRRCCCTCCTCCTRCCTCCR 116

Query Match   5.9%; Score 122.5; DB 2; Length 188;
Best Local Similarity   61.3%; Pred. No. 0.058;
Matches 19; Conservative   1; Mismatches   8; Indels   3; Gaps   2;

A;Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monocl
anes.
A;Reference number: A33625; MUID:90078352; PMID:2687294
A;Accession: B33625
A;Molecule type: protein
A;Residues: 1379-1384,'X',1386-1388,'X',1390-1398 <HE2>
A;Accession: A33625
A;Molecule type: protein
A;Residues: 1-21/Domain: signal sequence #status predicted <SIG>
F;22-4391/Product: perlecan #status predicted <MAT>
F;22-193/Domain: I <DOM1>
F;194-530/Domain: II <DOM2>
F;199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F;285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F;325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F;368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F;531-1677/Domain: III <DOM3>
F;1159-1206/Domain: laminin-type EGF-like homology <EG6>
F;1563-1610/Domain: laminin-type EGF-like homology <EG7>
F;1613-1668/Domain: laminin-type EGF-like homology <EG8>
F;1677-3686/Domain: IV <DOM4>
F;2007-2034/Domain: transmembrane #status predicted <TRM>
F;3845-4391/Domain: V <DOM5>
F;3888-3921/Domain: EGF homology <EGF1>
F;3995-4106/Domain: EGF homology <EGF>
F;4147-4175/Domain: EGF homology <EGF2>
F;4149-4151/Region: motor neuron attachment (L-R-E) motif
F;4299-4301/Region: motor neuron attachment (L-R-E) motif
F;65,71,76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
F;89,554,1755,2121,3072,3105,3279,3780,3836,4068/Binding site: carbohydrate (Asn) (co
F;2995,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

C;Superfamily: LDL receptor ligand-binding repeat homology; EGF homology; laminin G r
C;Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate; transmem
C;Cross-references: GDB:126372; OMIM:142461
A;Map position: 1p36.1-1p36.1
A;Gene: GDB:HSPG2
A;Genetics:

Query Match   5.9%; Score 122.5; DB 2; Length 4391;
Best Local Similarity   25.3%; Pred. No. 1.7;
Matches 62; Conservative 28; Mismatches 80; Indels 75; Gaps 12;

Qy   6 LTVPEAVGSG--SGNEVIE-GPQNATVLKGSQARNCTVSQG------------WKLINNALSDM 55
          ||| ||| | || | || ||| ||| |||||| || | | | || |||
Db   3007 VTVPPSEGSSYRLRSPVISIDPPSSTVQQCGQDASPKCLIHDGAAPISLEWKTRNQELEDN 3066

Qy   56 VVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC-------SLQ 105
          | | || | || || | | | ||| | || | | ||
Db   3067 VHIS--PNGSITT---------------IVGTRPSNHGTYRCVASNAYGVAQSVV 3104

Qy   106 NSRLHGSAYLTVQVMGELFI---PSVNL-VVAENEPCEYTCLPSHWTRLPDISWEL---- 157
          | | || | | ||| || | | ||||| | | | |||| | ||| |
Db   3105 NLSVHGPPTVSVLPEGPVWVGKAVTLECVSAGEPRS----SARWTRISSTPAKLEQRT 3160

Qy   158 -GLLVSHSSYFVPEPSDLQSAVSILALTPQSNGLITCVATWKSLKARKSATVNLTVIRC 216
          ||| |||| | |||| |||| | || || | | | || | || | | |||
Db   3161 YGLIMDSH---------------AVLQISSAKPSDAGTYVCLAQNALGTAQKQVEVIV---- 3202

Qy   217 PQDTG 221
          | ||
Db   3203 --DTG 3205

RESULT 15

JC1238

opioid-binding protein (clone DUZ1) - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 19-May-2000
C;Accession: JC1238
R;Lippman, D.A.; Lee, N.M.; Loh, H.H.

Gene 117, 249-254, 1992
A;Title: Opioid-binding cell adhesion molecule (OBCAM)-related clones from a rat brain c
A;Reference number: JC1238; MUID:92347701; PMID:1339369
A;Accession: JC1238
A;Molecule type: mRNA
A;Residues: 1-338 <LIP>
A;Cross-references: GB:M88709; NID:g203245; PIDN:AAA40858.1; PID:g203246
A;Experimental source: brain
C;Genetics:
A;Gene: OBCAM
C;Superfamily: carcinoembryonic antigen; carcinoembryonic antigen precursor amino-termin

Query Match   5.7%;   Score 118.5; DB 2; Length 338;
Best Local Similarity   25.2%; Pred. No. 0.22;
Matches   72; Conservative   36; Mismatches 107; Indels   71; Gaps   18;

Qy   5 LLTVPEAVGSGSGNEVI-EGPQNATVLKGSQARENCTVSOGWKLIMWALSDMVVL----- 58
          |!|:| |!: | !|:| | !|:| :| :|
Db   16 LLFIP-GVPVRSGDATFPKAMDNVTVRQGESATLRCTIDDRVTRVAW-LNRSTILYAGND 73

Qy   59 --SVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQ-----NSRLH 110
          |:| ! | ! |!:|
Db   74 KWSIDPRVILVN--TPTQY------SIMIQNVDVDEGPYTCSVQQTDNHPKTSRVH 122

Qy   111 GSAYLTVQVMGELFTPSVNLVVAENEPCEVTCL----PSHWTRLPDISWELGLLVSHSSY 166
          | /|!| |:| |:| |
Db   123 ----LIVQVPPQIMNISSDITVNEIISSVTLLCLAIGRPE-----PTVTWR----HLSV 167

Qy   167 Y----FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTV-----IRCP 217
          |: | !| :| !   :|
Db   168 KEGQGFVSEDEYLE----ISDIKRDQSGEYECSAL-NDVAAPDVRKVKITVNYPPYISKA 222

Qy   218 QDTGGGINIPGVL----SSLPSLGFSLPTWGK----VGLGLAGTML 255
          :|:| | |!| |!|
Db   223 KNTGVSVGQKGILSCEASAVPMAEFQ--WFKEDTRLATGLDGVRI 265

Search completed: April 28, 2003, 21:09:17
Job time : 27.0659 secs

OM protein - protein search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.4_p5_4578

Run on: April 29, 2003, 11:05:56 ; Search time 19.0659 Seconds
                 (without alignments)
                 1622.277 Million cell updates/sec

Title: US-09-729-264-6
Perfect score: 2077
Sequence: 1 MERHLLTVPEAVGSGSGCNEV.........HPQASFNLASPEKVSNTTVV 386

Scoring table: BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published_Applications_AA:*
                 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
                 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
                 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
                 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
                 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
                 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	145.5	7.0	405	1	US-08-755-235-4	Sequence 4, Appli
2	138.5	6.7	2473	9	US-10-184-644-559	Sequence 559, App
3	138.5	6.7	2473	9	US-10-184-634-559	Sequence 559, App
4	134	6.5	708	9	US-10-174-590-584	Sequence 584, App
5	134	6.5	708	9	US-10-176-758-584	Sequence 584, App
6	134	6.5	708	9	US-10-173-706-584	Sequence 584, App
7	134	6.5	708	9	US-10-173-705-584	Sequence 584, App
8	134	6.5	708	9	US-10-175-738-584	Sequence 584, App
9	134	6.5	708	9	US-10-175-752-584	Sequence 584, App
10	134	6.5	708	9	US-10-176-482-584	Sequence 584, App
11	134	6.5	708	9	US-10-176-757-584	Sequence 584, App
12	134	6.5	708	9	US-10-176-913-584	Sequence 584, App
13	134	6.5	708	9	US-10-180-552-584	Sequence 584, App
14	134	6.5	708	9	US-10-180-557-584	Sequence 584, App
15	134	6.5	708	9	US-10-173-700-584	Sequence 584, App
16	134	6.5	708	9	US-10-173-700-584	Sequence 584, App
17	134	6.5	708	9	US-10-174-572-584	Sequence 584, App
18	134	6.5	708	9	US-10-174-588-584	Sequence 584, App
19	134	6.5	708	9	US-10-174-588-584	Sequence 584, App
20	134	6.5	708	9	US-10-175-739-584	Sequence 584, App
21	134	6.5	708	9	US-10-175-740-584	Sequence 584, App
22	134	6.5	708	9	US-10-175-743-584	Sequence 584, App
23	134	6.5	708	9	US-10-176-488-584	Sequence 584, App
24	134	6.5	708	9	US-10-176-492-584	Sequence 584, App
25	134	6.5	708	9	US-10-176-747-584	Sequence 584, App
26	134	6.5	708	9	US-10-176-750-584	Sequence 584, App
27	134	6.5	708	9	US-10-176-985-584	Sequence 584, App
28	134	6.5	708	9	US-10-176-987-584	Sequence 584, App
29	134	6.5	708	9	US-10-176-991-584	Sequence 584, App
30	134	6.5	708	9	US-10-176-992-584	Sequence 584, App
31	134	6.5	708	9	US-10-176-993-584	Sequence 584, App
32	134	6.5	708	9	US-10-184-658-584	Sequence 584, App
33	134	6.5	708	9	US-10-173-695-584	Sequence 584, App
34	134	6.5	708	9	US-10-173-697-584	Sequence 584, App
35	134	6.5	708	9	US-10-173-705-584	Sequence 584, App
36	134	6.5	708	9	US-10-174-576-584	Sequence 584, App
37	134	6.5	708	9	US-10-174-585-584	Sequence 584, App
38	134	6.5	708	9	US-10-174-586-584	Sequence 584, App
39	134	6.5	708	9	US-10-175-747-584	Sequence 584, App
40	134	6.5	708	9	US-10-175-481-584	Sequence 584, App
41	134	6.5	708	9	US-10-176-481-584	Sequence 584, App
42	134	6.5	708	9	US-10-176-485-584	Sequence 584, App
43	134	6.5	708	9	US-10-176-487-584	Sequence 584, App
44	134	6.5	708	9	US-10-176-493-584	Sequence 584, App
45	134	6.5	708	9	US-10-176-911-584	Sequence 584, App

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
US-08-755-235-4

Sequence 4, Application US/08755235

Publication No. US2003005943A1

GENERAL INFORMATION:

APPLICANT: Stern, David M.
APPLICANT: Schmidt, Ann Marie
APPLICANT: Wu, Jun

TITLE OF INVENTION: METHOD FOR TREATING SYMPTOMS OF DIABETES

FILE REFERENCE: 0575/50159

CURRENT APPLICATION NUMBER: US/08/755,235

CURRENT FILING DATE: 1996-11-22

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patentin version 3.1

SEQ ID NO 4
LENGTH: 405
TYPE: PRT
ORGANISM: Human
US-08-755-235-4

Query Match   7.0%; Score 145.5; DB 1; Length 405;
Best Local Similarity   23.5%; Pred. No. 0.0034; Matches 38; Conservative 38; Mismatches 111; Indels 147; Gaps 20;

Qy   9 PEAVGSGSGNEVIEGPQNATVLKGSQARFNCTVSQG---WKLIMWALSDMVVLSVRPMEP 65
             | | | | | | | : || |: | | :|
Db   124 PEIVDSAS--ELITAGVPN----------------KVGTCVSEGSYPAGTLSWHLDG------KP 163

Qy   66 IITNDRFTS----QRYDQQGNFT--SEMIHNVEPSDSGNIR----CSLQNSRLHGSAY 114
             : ||| ||| || : | || || || | | |
Db   164 LVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRAL 220

Qy   115 LTVQVMGELFIP----SVNLVVAENEP--------CEVTCLPSHWTRLPDISWE 156
             | | | | | | || | | | || | | | | | |
Db   221 RTAPIQPRVWEPVPLEEVQLVV---EPEGGAVAPGGTVTLICEVPAQPS----PQIHWM 272

Qy   157 LGLLVSHSSSYYFVFEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKS-ATVNLTVIR 215
             | : | || | : | | :| | | | : | | |
Db   273 KD------GVPLPLPPSPVLILPEIGPQDQGTYSCVAPHSSHGPQESRAVVSISIIE 323

Qy   216 CPQDTGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLT-------PTCTLTIRCCC 268
             | | | | | | | | | | | | | | | | |
Db   -- 268

```
Db   324  -PGEEG------------------------PTAGSVGGSGLGTLALALGILGGGLGTRALLIGVIL  363
Qy   269  CRRCCGCNCCRCCFCCRRKRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQK  326
Db   364  WQR----------------QRRGEERKAPENQEEEERAELN------  390
Qy   327  TIETASLPPKSCESSDPEQRNSSCGPP  353
Db   391  ---------------QSEEPEAGESSTGGP  405
```

RESULT 2
US-10-184-644-559

; Sequence 559, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 559
; LENGTH: 2473
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-559

  Query Match   6.7%; Score 138.5; DB 9; Length 2473;
  Best Local Similarity   29.5%; Pred. No. 0.088;
  Matches   31; Conservative 3; Mismatches 36; Indels 35; Gaps 2;

```
Qy   183  ALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPT  242
Db  2274  AATTGAAGTTTCAATTAAAATTAATATGTTTCC---  2307
Qy   243  WGKVGLGLAGTMLLTPTCTLITIRCCCCR-RRCCGCNCCCRCCFCC  286
Db  2308  --------ATTCCTCATCGCCCACCCCACCCCGCGCCCCACCACC  2344
```

RESULT 3
US-10-184-634-559

; Sequence 559, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634

; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 559
; LENGTH: 2473
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-559

  Query Match   6.7%; Score 138.5; DB 9; Length 2473;
  Best Local Similarity   29.5%; Pred. No. 0.088;
  Matches   31; Conservative 3; Mismatches 36; Indels 35; Gaps 2;

```
Qy   183  ALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPT  242
Db  2274  AATTGAAGTTTCAATTAAAATTAATATGTTTCC---  2307
Qy   243  WGKVGLGLAGTMLLTPTCTLITIRCCCCR-RRCCGCNCCCRCCFCC  286
Db  2308  --------ATTCCTCATCGCCCACCCCACCCCGCGCCCCACCACC  2344
```

RESULT 4
US-10-174-590-584

; Sequence 584, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-174-590-584

  Query Match   6.5%; Score 134; DB 9; Length 708;
  Best Local Similarity   23.7%; Pred. No. 0.045;
  Matches   53; Conservative 40; Mismatches 101; Indels 30; Gaps 10;

```
Qy    14  SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT  73
Db    20  AGPSPHFLQOPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR------DLPG  71
Qy    74  SQRYDQGGNFTS--EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV-------QVMGE  122
Db    72  WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG  131
Qy   123  LFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ  176
Db   132  ---PSVSLVA--GVPANLICRSRGDARPPELLWFRDGVLDGATFHQTLLKEGTPGSVE  186
Qy   177  SAVSILALITPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT  220
Db   187  STLILTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT  230
```

RESULT 5

US-10-176-758-584
; Sequence 584, Application US/1017675B
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-584

Query Match   6.5%; Score 134; DB 9; Length 708;
Best Local Similarity 23.7%; Pred. No. 0.045;
Matches 53; Conservative 40; Mismatches 101; Indels 30; Gaps 10;

Qy   14 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
             ||||.|||::||:|| |||.|||:|||||||||||||||||||||||| :||| :|
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLAIGGQR-----------DLPG 71

Qy   74 SQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV-------QVMGE 122
            ||.|| || ||| ||| ||| |:| |:|
Db 132 ---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKEGTPGSVE 186

Qy 123 LFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ 176
            |||:|| | ||:||| ||:|:|| ::| |:| |:
Qy 177 SAVSILALTPQSNGTLTCVATWKSLKARKSAATVNLTVIRCPQDT 220
            | |:||| | ::| || :| || | |
Db 187 STLILTPFSHDDGATFVCRARSQALPTGRDTAITILSLQYPPEVT 230

RESULT 6
US-10-175-737-584
; Sequence 584, Application US/10175737
; Publication No. US20030013153A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C50
; CURRENT APPLICATION NUMBER: US/10/175,737
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 584

; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-737-584

Query Match   6.5%; Score 134; DB 9; Length 708;
Best Local Similarity 23.7%; Pred. No. 0.045;
Matches 53; Conservative 40; Mismatches 101; Indels 30; Gaps 10;

Qy   14 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
             ||||.|||::||:|| |||.|||:|||||||||||||||||||||||| :||| :|
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLAIGGQR-----------DLPG 71

Qy   74 SQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV-------QVMGE 122
            ||.|| || ||| ||| ||| |:| |:|
Db 132 ---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKEGTPGSVE 186

Qy 123 LFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ 176
            |||:|| | ||:||| ||:|:|| ::| |:| |:
Qy 177 SAVSILALTPQSNGTLTCVATWKSLKARKSAATVNLTVIRCPQDT 220
            | |:||| | ::| || :| || | |
Db 187 STLILTPFSHDDGATFVCRARSQALPTGRDTAITILSLQYPPEVT 230

RESULT 7
US-10-173-706-584
; Sequence 584, Application US/10173706
; Publication No. US20030022293A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C7
; CURRENT APPLICATION NUMBER: US/10/173,706
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-706-584

Query Match   6.5%; Score 134; DB 9; Length 708;
Best Local Similarity 23.7%; Pred. No. 0.045;
Matches 53; Conservative 40; Mismatches 101; Indels 30; Gaps 10;

Qy   14 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
             ||||.|||::||:|| |||.|||:|||||||||||||||||||||||| :||| :|
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLAIGGQR-----------DLPG 71

Qy   74 SQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV-------QVMGE 122
            ||.|| || ||| ||| ||| |:| |:|
Db 132 ---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKEGTPGSVE 186

Qy 123 LFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ 176
            |||:|| | ||:||| ||:|:|| ::| |:| |:
Qy 177 SAVSILALTPQSNGTLTCVATWKSLKARKSAATVNLTVIRCPQDT 220
            | |:||| | ::| || :| || | |
Db 187 STLILTPFSHDDGATFVCRARSQALPTGRDTAITILSLQYPPEVT 230

RESULT 8

```
US-10-175-738-584
; Sequence 584, Application US/10175738
; Publication No. US20030022294A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT: Chen,Jian
;  APPLICANT: Desnoyers,Luc
;  APPLICANT: Goddard,Audrey
;  APPLICANT: Pan,James
;  APPLICANT: Gurney,Austin L.
;  APPLICANT: Godowski,Paul J.
;  APPLICANT: Watanabe,Colin K.
;  APPLICANT: Wood,William I.
;  APPLICANT: Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION: ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C45
;  CURRENT APPLICATION NUMBER: US/10/175,738
;  CURRENT FILING DATE: 2002-06-19
;  Prior application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
;  SEQ ID NO 584
;  LENGTH: 708
;  TYPE: PRT
;  ORGANISM: Homo Sapien
US-10-175-738-584

Query Match   6.5%;  Score 134;  DB 9;  Length 708;
Best Local Similarity  23.7%;  Pred. No. 0.045;
Matches  53;  Conservative  40;  Mismatches 101;  Indels  30;  Gaps  10;

Qy   14 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
         ::  ::  | |||| ||| || ||  |  || ||  ::
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR--------DLPG 71

Qy   74 SQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV--------QVMGE 122
         ||          |:| ||| ||   |  |  ||    |:|          | |
Db   72 WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy  123 LFIPSVNLVVAAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ 176
         |||:||   |  | |    | | ::|  ::|: ::| ||  |  :  |  | ::
Db  132 ---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKEGTPGSVE 186

Qy  177 SAVSILALTPQSNGTLITCVATWKSLKARKSATVNLTVIRCPQDT 220
         | ||:  | |  | | | ||  :: |   | :||:  |: | |
Db  187 STLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 9

```
US-10-175-752-584
; Sequence 584, Application US/10175752
; Publication No. US20030022295A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT: Chen,Jian
;  APPLICANT: Desnoyers,Luc
;  APPLICANT: Goddard,Audrey
;  APPLICANT: Pan,James
;  APPLICANT: Gurney,Austin L.
;  APPLICANT: Godowski,Paul J.
;  APPLICANT: Smith,Victoria
;  APPLICANT: Watanabe,Colin K.
;  APPLICANT: Wood,William I.
;  APPLICANT: Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION: ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C60
;  CURRENT APPLICATION NUMBER: US/10/175,752
```

RESULT 10

```
US-10-176-482-584
; Sequence 584, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:
;  APPLICANT: Baker,Kevin P.
;  APPLICANT: Chen,Jian
;  APPLICANT: Desnoyers,Luc
;  APPLICANT: Goddard,Audrey
;  APPLICANT: Pan,James
;  APPLICANT: Gurney,Austin L.
;  APPLICANT: Godowski,Paul J.
;  APPLICANT: Smith,Victoria
;  APPLICANT: Watanabe,Colin K.
;  APPLICANT: Wood,William I.
;  APPLICANT: Zhang,Zemin
;  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;  TITLE OF INVENTION: ACIDS ENCODING THE SAME
;  FILE REFERENCE: P3430R1C70
;  CURRENT APPLICATION NUMBER: US/10/176,482
;  CURRENT FILING DATE: 2002-06-20
;  Prior Application removed - See File Wrapper or Palm
;  NUMBER OF SEQ ID NOS: 612
;  SEQ ID NO 584
;  LENGTH: 708
;  TYPE: PRT
;  ORGANISM: Homo Sapien
US-10-176-482-584

Query Match   6.5%;  Score 134;  DB 9;  Length 708;
Best Local Similarity  23.7%;  Pred. No. 0.045;
Matches  53;  Conservative  40;  Mismatches 101;  Indels  30;  Gaps  10;

Qy   14 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
         ::  ::  | |||| ||| || ||  |  || ||  ::
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR--------DLPG 71

Qy   74 SQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV--------QVMGE 122
         ||          |:| ||| ||   |  |  ||    |:|          | |
Db   72 WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy  123 LFIPSVNLVVAAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ 176
         |||:||   |  | |    | | ::|  ::|: ::| ||  |  :  |  | ::
Db  132 ---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKEGTPGSVE 186

Qy  177 SAVSILALTPQSNGTLITCVATWKSLKARKSATVNLTVIRCPQDT 220
         | ||:  | |  | | | ||  :: |   | :||:  |: | |
Db  187 STLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 11
US-10-176-757-584
Sequence 584, Application US/10176757
Publication No. US20030022297A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Chen,Jian
APPLICANT: Desnoyers,Luc
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C86
CURRENT APPLICATION NUMBER: US/10/176,757
CURRENT FILING DATE: 2002-06-20
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
LENGTH: 708
TYPE: PRT
ORGANISM: Homo Sapien
US-10-176-757-584

Query Match   6.5%;   Score 134;   DB 9;   Length 708;
Best Local Similarity   23.7%;   Pred. No. 0.045;
Matches 53;   Conservative 40;   Mismatches 101;   Indels 30;   Gaps 10;

```
Qy   14 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR------------ 71

Qy   74 SQRYDQGGNFTS---EMIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV------QVMGE 122
Db   72 WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy  123 LFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ 176
Db  132 ---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKEGTPGSVE 186

Qy  177 SAVSILALPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 220
Db  187 STLILTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 12
US-10-176-913-584
Sequence 584, Application US/10176913
Publication No. US20030022298A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Chen,Jian
APPLICANT: Desnoyers,Luc
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C66
CURRENT APPLICATION NUMBER: US/10/176,913
CURRENT FILING DATE: 2002-06-20
Prior Application removed - See file Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
LENGTH: 708
TYPE: PRT
ORGANISM: Homo Sapien
US-10-176-913-584

Query Match   6.5%;   Score 134;   DB 9;   Length 708;
Best Local Similarity   23.7%;   Pred. No. 0.045;
Matches 53;   Conservative 40;   Mismatches 101;   Indels 30;   Gaps 10;

```
Qy   14 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR------------ 71

Qy   74 SQRYDQGGNFTS---EMIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV------QVMGE 122
Db   72 WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy  123 LFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ 176
Db  132 ---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKEGTPGSVE 186

Qy  177 SAVSILALPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 220
Db  187 STLILTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 13
US-10-180-552-584
Sequence 584, Application US/10180552
Publication No. US20030022300A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Chen,Jian
APPLICANT: Desnoyers,Luc
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C153
CURRENT APPLICATION NUMBER: US/10/180,552
CURRENT FILING DATE: 2002-06-25
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
LENGTH: 708
TYPE: PRT
ORGANISM: Homo Sapien
US-10-180-552-584

Query Match   6.5%;   Score 134;   DB 9;   Length 708;
Best Local Similarity   23.7%;   Pred. No. 0.045;
Matches 53;   Conservative 40;   Mismatches 101;   Indels 30;   Gaps 10;

```
Qy   14 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR------------ 71

Qy   74 SQRYDQGGNFTS---EMIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV------QVMGE 122
```

```
Db  72 WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy 123 LFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ 176
Db 132 ---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKEGTPGSVE 186

Qy 177 SAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 220
Db 187 STILTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 14
US-10-180-557-584
Sequence 584, Application US/10180557
Publication No. US20030223019A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C147
CURRENT APPLICATION NUMBER: US/10/180,557
CURRENT FILING DATE: 2002-06-25
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
  LENGTH: 708
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-180-557-584

Query Match   6.5%; Score 134; DB 9; Length 708;
Best Local Similarity   23.7%; Pred. No. 0.045;
Matches   53; Conservative   40; Mismatches 101; Indels   30; Gaps   10;

```
Qy  14 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
Db  20 AGPSPHFLQQPEDLVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR--------DLPG 71

Qy  74 SQRYDQGGNFTS----EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLIV------QVMGE 122
Db  72 WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy 123 LFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ 176
Db 132 ---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKEGTPGSVE 186

Qy 177 SAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 220
Db 187 STILTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 15
US-10-173-700-584
Sequence 584, Application US/10173700
Publication No. US20030027262A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C14
CURRENT APPLICATION NUMBER: US/10/173,700
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
  LENGTH: 708
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-173-700-584

Query Match   6.5%; Score 134; DB 9; Length 708;
Best Local Similarity   23.7%; Pred. No. 0.045;
Matches   53; Conservative   40; Mismatches 101; Indels   30; Gaps   10;

```
Qy  14 SGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
Db  20 AGPSPHFLQQPEDLVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR--------DLPG 71

Qy  74 SQRYDQGGNFTS----EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLIV------QVMGE 122
Db  72 WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy 123 LFIPSVNLVVAENEPCEVTCLPSHWTR-LPDISW-ELGLLVSHSSYY--FVPE--PSDLQ 176
Db 132 ---PSVSLVA--GVPANLTCRSRGDARPTPELLWFRDGVLLDGATFHQTLLKEGTPGSVE 186

Qy 177 SAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 220
Db 187 STILTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

Search completed: April 29, 2003, 20:17:34
Job time : 27.0659 secs

OM protein - protein search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 18:14:43 ; Search time 14.0485 Seconds
                   (without alignments)
                   808.430 Million cell updates/sec

Title: US-09-729-264-6
Perfect score: 2077
Sequence: 1 MERHLLTVPEAVGSGSGNEV..........HPQASFNLASPEKVSNTTVV 386
Scoring table: BLOSUM62
                   Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_AA:*
            1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
            2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
            3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
            4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
            5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
            6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	128.5	6.2	1345	2	US-08-977-767-3	Sequence 3, Appli
2	128	6.2	1447	4	US-09-041-886-25	Sequence 25, Appl
3	128	6.2	1447	5	PCT-US94-05277-2	Sequence 2, Appli
4	127.5	6.1	869	4	US-08-374-834-16	Sequence 16, Appl
5	127.5	6.1	869	2	US-08-644-271-29	Sequence 29, Appl
6	127.5	6.1	869	4	US-09-077-955-33	Sequence 33, Appl
7	127	6.1	332	4	US-09-062-365-1	Sequence 1, Appli
8	124.5	6.0	340	4	US-09-651-200-2	Sequence 2, Appli
9	124.5	6.0	441	4	US-09-651-200-4	Sequence 4, Appli
10	123.5	5.9	534	4	US-09-651-200-6	Sequence 6, Appli
11	123.5	5.9	534	4	US-09-651-200-24	Sequence 24, Appl
12	122	5.9	318	2	US-08-633-148-4	Sequence 4, Appli
13	122	5.9	340	2	US-08-633-148-2	Sequence 2, Appli
14	114	5.5	868	2	US-08-374-834-1	Sequence 1, Appli
15	114	5.5	868	1	US-08-644-271-1	Sequence 1, Appli
16	114	5.5	868	4	US-09-077-955-1	Sequence 1, Appli
17	113	5.4	1395	4	US-09-540-245A-15	Sequence 15, Appl
18	111	5.3	689	4	US-09-499-964-1	Sequence 1, Appli
19	111	5.3	365	2	US-08-979-424-3	Sequence 3, Appli
20	111	5.3	365	4	US-09-272-496-2	Sequence 2, Appli
21	109.5	5.3	325	4	US-09-651-200-20	Sequence 20, Appl
22	109.5	5.3	478	5	PCT-US99-08493-15	Sequence 15, Appl
23	109.5	5.3	860	5	PCT-US95-08493-19	Sequence 19, Appl
24	109.5	5.3	868	5	PCT-US95-08493-21	Sequence 21, Appl
25	109.5	5.3	890	1	US-08-445-640-2	Sequence 2, Appli
26	109.5	5.3	890	3	US-08-170-558-2	Sequence 2, Appli
27	109.5	5.3	890	3	US-08-447-314-2	Sequence 2, Appli
28	109.5	5.3	890	3	US-08-445-461-2	Sequence 2, Appli
29	109.5	5.3	946	5	US-08-286-305A-1	Sequence 1, Appli
30	109.5	5.3	911	1	US-08-441-104A-1	Sequence 1, Appli
31	109.5	5.3	911	2	US-08-440-816A-1	Sequence 1, Appli
32	109.5	5.3	911	4	US-09-417-381A-1	Sequence 1, Appli
33	107.5	5.2	362	1	US-08-415-751-6	Sequence 6, Appli
34	106.5	5.1	946	5	PCT-US95-08493-13	Sequence 13, Appl
35	106	5.1	801	3	US-07-906-349A-6	Sequence 6, Appli
36	105	5.1	319	1	US-08-597-495B-22	Sequence 22, Appl
37	105	5.1	319	4	US-08-068-051A-22	Sequence 22, Appl
38	105	5.1	319	4	US-09-336-536-67	Sequence 67, Appl
39	105	5.1	319	4	US-09-254-465A-6	Sequence 6, Appli
40	105	5.1	690	4	US-08-928-383B-2	Sequence 2, Appli
41	105	5.1	690	4	US-09-553-433-2	Sequence 2, Appli
42	105	5.1	690	4	US-08-935-433-2	Sequence 2, Appli
43	105	5.1	1297	4	US-09-540-245A-17	Sequence 17, Appl
44	105	5.1	1381	4	US-09-540-245A-16	Sequence 16, Appl
45	104.5	5.0	879	1	US-08-554-612C-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-977-767-3
Sequence 3, Application US/08977767
Patent No. 5972684
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Yue, Henry
APPLICANT: Greenwald, Sara
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: CARBONIC ANHYDRASE VIII
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/977,767
FILING DATE: Herewith
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0423 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1345 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1532042
US-08-977-767-3

Query Match   6.2%; Score 128.5; DB 2; Length 1345;

Left column

```
Qy  190 GTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGI----NIPGVLSLSLPSLGFSLPTWGK  245
        |||                           ||||
Db  414 GTCTCTGT--------------GC-CGTGGGAAGCGTCAGAGCCCGTGGATGTGGA  455

Qy  246 VGLGLAGTMLLT-PTCTLTIRCCCRRRCGCGNCCCRC---------------CFCC  286
                 ||||  |||                                  ||
Db  456 CGTGAAGAGGGGTTCCTCTATGACCCCTTCCTGCCCCTTCTGAGACTCAGCACC  507
```

Best Local Similarity 33.0%; Pred. No. 0.0098; Indels 33; Gaps   5;
Matches 37; Conservative   1; Mismatches 41;

RESULT 2
US-09-041-886-25
Patent No. 6235872
Sequence 25, Application US/09041886
GENERAL INFORMATION:
    APPLICANT: Bredesen, Dale E.
    APPLICANT: Rabizadeh, Sharroz
    TITLE OF INVENTION: Proapoptotic Peptides, Dependence
    TITLE OF INVENTION: Polypeptides and Methods of Use
    NUMBER OF SEQUENCES: 72
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Campbell & Flores LLP
        STREET: 4370 La Jolla Village Drive, Suite 700
        CITY: San Diego
        STATE: California
        COUNTRY: United States
        ZIP: 92122
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/09/041,886
        FILING DATE:
        CLASSIFICATION:
    ATTORNEY/AGENT INFORMATION:
        NAME: Campbell, Cathryn A.
        REGISTRATION NUMBER: 31,815
        REFERENCE/DOCKET NUMBER: P-LJ 2626
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (619) 535-9001
        TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 25:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 1447 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
    MOLECULE TYPE: protein
US-09-041-886-25

Query Match   6.2%; Score 128; DB 4; Length 1447;
Best Local Similarity   24.8%; Pred. No. 0.012;
Matches 60; Conservative 33; Mismatches 99; Indels 50; Gaps 10;

Qy   14 SGSGNEV--------------------IEGPQNATVLKGSQARFNCTVSQGW--KLIMWALSDM   55
           :||| ::|:|:| :   ::| |::||
Db 220 SRTGNEAEVRILSDPGLHRQLYFLQRPSNVVAIEGKDAVLECCVS-GYPPPSFTWLRGEE 278

Qy   56 VVLSVRPMEPIIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSA 113
Db 279 VI-------------QLRSKYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNENISASA 323

Qy 114 YLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISW-ELGLLVSHSSSYYFVPEP 172
Db 324 ELTVLVPPWFLNHPSNLYAYESMDIEFECTVS-GKPVPTVNWMKNGDVVIPSDYFQIVGG 382

Qy 173 SDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSS 232
Db 383 SNLR----ILGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS------VLPS 430
```

Right column

```
Qy  233 LP 234
        |
Db  431 AP 432
```

RESULT 3
PCT-US94-05277-2
Sequence 2, Application PC/TUS9405277
GENERAL INFORMATION:
    APPLICANT: Bruskin, Arthur
    APPLICANT: Jarosz, David E.
    APPLICANT: Johnson, Karen
    APPLICANT: Kinzler, Kenneth W.
    APPLICANT: Vogelstein, Bert
    APPLICANT: Zabrecky, James R.
    NUMBER OF SEQUENCES: 2
    TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Banner, Birch, McKie & Beckett
        STREET: 1001 G Street, N.W.
        CITY: Washington
        STATE: D.C.
        COUNTRY: USA
        ZIP: 20001
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: PatentIn Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: PCT/US94/05277
        FILING DATE:
        CLASSIFICATION:
    ATTORNEY/AGENT INFORMATION:
        NAME: Kagan, Sarah A.
        REGISTRATION NUMBER: 32,141
        REFERENCE/DOCKET NUMBER: 01107.42709
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 202.508.9100
        TELEFAX: 202.508.9299
        TELEX: 197430 BBMB UT
INFORMATION FOR SEQ ID NO: 2:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 1447 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
    MOLECULE TYPE: protein
PCT-US94-05277-2

Query Match   6.2%; Score 128; DB 5; Length 1447;
Best Local Similarity   24.8%; Pred. No. 0.012;
Matches 60; Conservative 33; Mismatches 99; Indels 50; Gaps 10;

Qy   14 SGSGNEV--------------------IEGPQNATVLKGSQARFNCTVSQGW--KLIMWALSDM   55
           :||| ::|:|:| :   ::| |::||
Db 220 SRTGNEAEVRILSDPGLHRQLYFLQRPSNVVAIEGKDAVLECCVS-GYPPPSFTWLRGEE 278

Qy   56 VVLSVRPMEPIIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSRLHGSA 113
Db 279 VI-------------QLRSKYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNENISASA 323

Qy 114 YLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISW-ELGLLVSHSSSYYFVPEP 172
Db 324 ELTVLVPPWFLNHPSNLYAYESMDIEFECTVS-GKPVPTVNWMKNGDVVIPSDYFQIVGG 382

Qy 173 SDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSS 232
Db 383 SNLR----ILGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS------VLPS 430

Qy 233 LP 234
        |
Db 431 AP 432
```

RESULT 4
US-08-374-834-16
Sequence 16, Application US/08374834
Patent No. 5656473
GENERAL INFORMATION:
APPLICANT: Valenzuela, et al.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTOR
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US 08/374,834
FILING DATE: 19-JAN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/095,658
FILING DATE: 21-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J.
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 190A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 345-7400
TELEFAX: (914) 345-7721
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 869 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-374-834-16

Query Match     6.1%;   Score 127.5;  DB 1;  Length 869;
Best Local Similarity 20.9%;  Pred. No. 0.0065;
Matches 86; Conservative 51; Mismatches 176;  Indels 99;  Gaps 19;

```
Qy   19 EVIEGPQNATVLKGSQOARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY 77
           ::   :|  |: :||  |||
Db  122 KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD-------SPLRENSRIAVLE- 171

Qy   78 DQGGNPTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMGELFIPSVNLVVAENE 136
         | |   | : |
Db  172 -SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKVVKLEVEVFARILRAPESHNV 223

Qy  137 P--------CEVTCLPSHWTRLPDISW-ELGLLVSHSSYFVPEPSDLQSAVSILALTPQ 187
         |        |
Db  224 TFGSEVTLHCTATGIP-----VPTIWIENGNAVSSGSIQESVKDRVIDSRLQLFITKP- 277

Qy  188 SNGTLTCVAT---WKSLKARKSATVNLTVIRCPQDTGGG-------INIPGVLSSLP 234
           | ||:|| :|  : | | |
Db  278 --GLYICIATNKHGEKFSTAKAAATISIAEWSKPQKDNKGYCAQYRGEVCNAVLAKDALV 335

Qy  235 SLGFSL----------PTWGKVGLGLAGTMLLTPTCTLTIRCCCCRRRCCGCN---- 277
         ||:||           : |
Db  336 FLNTSTADPEEAQELLVHTAWNEL-----KVVSPVCRPAAEALLCNHIFQECSPGVVP 388

Qy  278 ----CCCRCC-------FCCRRKRGFRIQFQKKSEKETNKETETESGNENSGYNSDEQKT 327
             |||||         : |: |
Db  389 TPIPICREYCLAVKELPCAKE---WLVMEEKTHRGLYRSEMHLLSVPECSKLPSMHWDP 444

Qy  328 TETASLPPKSCESSDPEQRNSSCGPPHQRADQRP---PRPASHPQASFNLA 375
```

Db 445 TACARLP-----HLDYNKENLKTFPP--MTSSKPSVDIPNLPSSSSSFSVS 489

RESULT 5
US-08-644-271-29
Sequence 29, Application US/08644271
Patent No. 5814478
GENERAL INFORMATION:
APPLICANT: Valenzuela, et al.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
TITLE OF INVENTION: AND LIGANDS
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill Road
CITY: Tarrytown
COUNTRY: USA
STATE: NY
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/644,271
FILING DATE: 10-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/008,657
FILING DATE: 15-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Cobert, Robert J
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 195A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
TELEX:
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 869 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-644-271-29

Query Match   6.1%;   Score 127.5; DB 2; Length 869;
Best Local Similarity 20.9%; Pred. No. 0.0065;
Matches 86; Conservative 51; Mismatches 176; Indels 99; Gaps 19;

```
Qy   19 EVIEGPQNATVLKGSQOARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY 77
           ::   :|  |: :||  |||
Db  122 KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD-------SPLRENSRIAVLE- 171

Qy   78 DQGGNPTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMGELFIPSVNLVVAENE 136
         | |   | : |
Db  172 -SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKVVKLEVEVFARILRAPESHNV 223

Qy  137 P--------CEVTCLPSHWTRLPDISW-ELGLLVSHSSYFVPEPSDLQSAVSILALTPQ 187
         |        |
Db  224 TFGSEVTLHCTATGIP-----VPTIWIENGNAVSSGSIQESVKDRVIDSRLQLFITKP- 277

Qy  188 SNGTLTCVAT---WKSLKARKSATVNLTVIRCPQDTGGG-------INIPGVLSSLP 234
           | ||:|| :|  : | | |
Db  278 --GLYICIATNKHGEKFSTAKAAATISIAEWSKPQKDNKGYCAQYRGEVCNAVLAKDALV 335

Qy  235 SLGFSL----------PTWGKVGLGLAGTMLLTPTCTLTIRCCCCRRRCCGCN---- 277
         ||:||           : |
Db  336 FLNTSTADPEEAQELLVHTAWNEL-----KVVSPVCRPAAEALLCNHIFQECSPGVVP 388
```

RESULT 6
US-09-077-955-33
Patent No. 6413740
GENERAL INFORMATION:
APPLICANT: Valenzuela et al., David M.
TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
FILE REFERENCE: REG195-B-PCT-US
CURRENT APPLICATION NUMBER: US/09/077,955A
CURRENT FILING DATE: 1998-09-10
EARLIER APPLICATION NUMBER: PCT/US96/20696
EARLIER FILING DATE: 1996-12-13
EARLIER APPLICATION NUMBER: 08/644,271
EARLIER FILING DATE: 1996-05-10
EARLIER APPLICATION NUMBER: 60/008,657
EARLIER FILING DATE: 1995-12-15
NUMBER OF SEQ ID NOS: 36
SEQ ID NO 33
TYPE: PRT
ORGANISM: Homo sapiens
SOFTWARE: PatentIn Ver. 2.0
LENGTH: 869
US-09-077-955-33

Query Match   6.1%;   Score 127.5; DB 4; Length 869;
Best Local Similarity   20.9%; Pred. No. 0.0065;
Matches 86; Conservative 51; Mismatches 176; Indels 99; Gaps 19;

```
Qy   19 EVIEGPQNATVLKGSQARFNCTVSQGWK-LIMWALSDMVLSVRPMEPIITNDRFTSQRY 77
Db  122 KITRPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD------------SPLRENSRIAVLE- 171

Qy   78 DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LIVQVMGELFIPSVNLVVAENE 136
Db  172 -----SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKVVKLEVEVFARILRAPESHNV 223

Qy  137 P-------CEVTCLPSHWTRLPDISW-ELGLLVSHSSYFVPEPSDLQSAVSILALTPQ 187
Db  224 TFGSFVTLHCTVATGIP-----VPTITWIENGNAVSSGSIQESVKDRVIDSRLQLFITKP- 277

Qy  188 SNGTLTCVAI-----WKSLKARKSATVNLTVIRCPQDTGGG--------INIPGVLSSLP 234
Db  278 --GLYTCIATNKHGEKFSTAKAAAITSIAEWSKPQKDNKGYCAQYRGEVCNAVLAKDALV 335

Qy  235 SLGFSL------PTWGKVGIGLAGTMLLIPTCLTLTIRCCCCRRCCGCN----- 277
Db  336 FLNTSYADPEEAQELLVHTAWNEL----KVVSPVCRPAAEALLCNHFQECSPGVVP 388

Qy  278 -----CCCRCC----------FCCRRKRGFRIQFQKKSEKEKTNKEIETESGNENSGYNSDEQKT 327
Db  389 TPIPICREYCLAVKELFCAKE----WLVMEEKTHRGLYRSEMHLLSVPECSKLPSMHWDP 444

Qy  328 TETASLPPKSCESSDPEQRNSSCGPPHQRADQRP----PRPASHPQASFNLA 375
Db  445 TACARLP-----HLDYNKENLKTFPP-----MTSSKPSVDIPNLPSSSSSFSVS 489
```

TITLE OF INVENTION: METHOD FOR INHIBITING TUMOR INVASION OR SPREADING IN A
TITLE OF INVENTION: SUBJECT
FILE REFERENCE: 55424
CURRENT APPLICATION NUMBER:
CURRENT FILING DATE: 1998-04-17
NUMBER OF SEQ ID NOS: 6
SEQ ID NO 1
LENGTH: 332
TYPE: PRT
ORGANISM: Human
US-09-062-365-1

Query Match   6.1%;   Score 127; DB 4; Length 332;
Best Local Similarity   25.0%; Pred. No. 0.0019;
Matches 70; Conservative 31; Mismatches 87; Indels 15; Gaps 15;

```
Qy    9 PEAVGSGSGNEVIEGPQNATVLKGSQARFNCTVSQG---WKLIMWALSDMVLSVRPMEP 65
Db  102 PEIVDSAS--ELTAGVPN---------KVGTCVSEGSYPAGTLSWHLDG-------KP 141

Qy   66 IITNDRFTS-----QRYDQGGNFT--SEMIIHNVEPSDSGNIR-----CSLQNSRLHGSAY 114
Db  142 LVPNEKGVSVKEQTRRHPETGLFTLQSELM--VTPARGDPRPTFSCSFSPGLPRHRAL 198

Qy  115 LIVQVMGELFIP----SVNLVVAENEP------CEVTCLPSHWTRLPDISWE 156
Db  199 RTAPTQPRVWBPVPLEEVQLVV--EPEGGAVAPGGTVTLTCEVPAQPS----PQIHWM 250

Qy  157 LGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRC 216
Db  251 KD------GVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIE- 300

Qy  217 PQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLL 256
Db  301 PGEEG------------------PTAGSVGGSGLGTLAL 321
```

RESULT 7
US-09-062-365-1
Patent No. 6465422
Sequence 1, Application US/09062365
GENERAL INFORMATION:
APPLICANT: Schmidt, Ann Marie
APPLICANT: Stern, David

RESULT 8
US-09-651-200-2
Patent No. 6429303
Sequence 2, Application US/09651200
GENERAL INFORMATION:
APPLICANT: Green et al
TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
TITLE OF INVENTION: Polypeptides Encoded Thereby
FILE REFERENCE: 15966-562 (CURA-62)
CURRENT APPLICATION NUMBER: US/09/651,200
CURRENT FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/152383
PRIOR FILING DATE: 1999-09-03
PRIOR APPLICATION NUMBER: 60/172909
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/183578
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 340
TYPE: PRT
ORGANISM: Homo sapiens
US-09-651-200-2

Query Match   6.0%;   Score 124.5; DB 4; Length 340;
Best Local Similarity   21.2%; Pred. No. 0.0033;
Matches 77; Conservative 51; Mismatches 127; Indels 109; Gaps 17;

```
Qy   14 SGSGNEVIEGPQNATV-LKGSQARFNCTVS--QGWKL----IMWALSDMVLSVRPMEPI 66
Db   48 SPTGAVEVQVPEDVVALVGTDATLHCSPSPEPGFSLTQLNLIWQLTDTKQLV------- 100

Qy   67 ITNDRFTSQRYDQGGNF---------------TSEMIIHNVEPSDSGNIRCSLQNSRLHGS 112
```

```
Db    101  ---HSFTEGR-DQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFV-SIRDFGS 155
           ||||         | ||  | | :|:|||  |:||
Qy    113  AYLTVQVMGELFIPSVNLV----VAENEPCEVTCLPSHWTRLP--DISWELGL---LVSH 163
           |||:| |            : ::|||         || |:||
Db    156  AAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPLTGN 213
           |:| |||:|| :     :|                     | | :|
Qy    164  SSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP---QDT 220
                  :|::  || |  |                              :|:| |  |:
Db    214  VVTSQMANEQGLFDVHSVLRVVLGANGTYSC------------------LVRNPVLQQDA 255

Qy    221  GGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCRRRCCGCNCCC 280
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    256  HGSVTITGQPMTFPPEAL----WTVGLSVCLIALLV--------------------- 288

Qy    281  RCCFCCCRRKRGFRIQFQKKSEKEKTNKEFETESGNENSGYNSDEQKTTETASLPPKSCES 340
           | ::| | |  :::|| |:||||     |||||||||||| ||| :|
Db    289  ALAFVCWRK-------------------IKQSCEEENAGAEDQDG--------EGEGSKTALQPLKHSDS 331

Qy    341  SDPE 344
           : |
Db    332  KEDD 335
```

```
RESULT 9
US-09-651-200-4
; Sequence 4, Application US/09651200
; Patent No. 6429303
; GENERAL INFORMATION:
; APPLICANT: Green et al
; TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
; TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: 15966-562 (CURA-62)
; CURRENT APPLICATION NUMBER: US/09/651,200
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/152383
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/172909
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/183578
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 4
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-651-200-4

Query Match              6.0%;  Score 124.5;  DB 4;  Length 441;
Best Local Similarity   21.2%;  Pred. No. 0.0047;
Matches     77;  Conservative   51;  Mismatches 127;  Indels 109;  Gaps  17;

Qy     14  SGSGNEVIEGPQNATV-LKGSQARFNCTVS--QGWKL----IWWALSDMVVLSVRPMEPI 66
           |  :|             ::: ||             |:||
Db    149  SPTGAVEVQVPEDPVVALVGTDATFLHCSFSPEPGFSLTQLNLIWQLTDTKQLV 201

Qy     67  ITNDRFTSQRYDQGGNF--------------TSEMIIHNVEPSDSGNIRCSLQNSRLHGS 112
              |: | | |                   : :| ::||
Db    202  --HSFTEGR-DQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFV-SIRDFGS 256

Qy    113  AYLTVQVMGELFIPSVNLV----VAENEPCEVTCLPSHWTRLP--DISWELGL---LVSH 163

Db    257  AAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPLTGN 314

Qy    164  SSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP--QDT 220

Db    315  VVTSQMANEQGLFDVHSVLRVVLGANGTYSC------------------LVRNPVLQQDA 356

Qy    221  GGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCRRRCCGCNCCC 280

Db    357  HGSVTITGQPMTFPPEAL----WTVGLSVCLIALLV--------------------- 389
```

```
Qy    281  RCCFCCCRRKRGFRIQFQKKSEKEKTNKEFETESGNENSGYNSDEQKTTETASLPPKSCES 340
           | ::| | |  :::|| |:||||     |||||||||||| ||| :|
Db    390  ALAFVCWRK-------------------IKQSCEEENAGAEDQDG--------EGEGSKTALQPLKHSDS 432

Qy    341  SDPE 344
           : |
Db    433  KEDD 436
```

```
RESULT 10
US-09-651-200-6
; Sequence 6, Application US/09651200
; Patent No. 6429303
; GENERAL INFORMATION:
; APPLICANT: Green et al
; TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
; TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
; TITLE OF INVENTION: Polypeptides Encoded Thereby
; FILE REFERENCE: 15966-562 (CURA-62)
; CURRENT APPLICATION NUMBER: US/09/651,200
; CURRENT FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: 60/152383
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/172909
; PRIOR FILING DATE: 1999-12-21
; PRIOR APPLICATION NUMBER: 60/183578
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
;
; SEQ ID NO 6
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-651-200-6

Query Match              5.9%;  Score 123.5;  DB 4;  Length 534;
Best Local Similarity   21.2%;  Pred. No. 0.0076;
Matches     77;  Conservative   50;  Mismatches 128;  Indels 109;  Gaps  17;

Qy     14  SGSGNEVIEGPQNATV-LKGSQARFNCTVS--QGWKL----IWWALSDMVVLSVRPMEPI 66
           |  :|             ::: ||             |:||
Db    242  SPTGAVEVQVPEDPVVALVGTDATFLRCSFSPEPGFSLAQLNLIWQLTDTKQLV------ 294

Qy     67  ITNDRFTSQRYDQGGNF--------------TSEMIIHNVEPSDSGNIRCSLQNSRLHGS 112

Db    295  ---HSFTEGR-DQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFV-SIRDFGS 349

Qy    113  AYLTVQVMGELFIPSVNLV----VAENEPCEVTCLPSHWTRLP--DISWELGL---LVSH 163

Db    350  AAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPLTGN 407

Qy    164  SSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP---QDT 220

Db    408  VVTSQMANEQGLFDVHSVLRVVLGANGTYSC------------------LVRNPVLQQDA 449

Qy    221  GGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCRRRCCGCNCCC 280

Db    450  HGSVTITGQPMTFPPEAL----WTVGLSVCLIALLV--------------------- 482

Qy    281  RCCFCCCRRKRGFRIQFQKKSEKEKTNKEFETESGNENSGYNSDEQKTTETASLPPKSCES 340

Db    483  ALAFVCWRK-------------------IKQSCEEENAGAEDQDG--------EGEGSKTALQPLKHSDS 525

Qy    341  SDPE 344

Db    526  KEDD 529
```

```
RESULT 11
US-09-651-200-24
; Sequence 24, Application US/09651200
; Patent No. 6429303
```

GENERAL INFORMATION:

~ ; APPLICANT: Green et al
~ ; TITLE OF INVENTION: Polynucleotides Encoding Members of the Human B
~ ; TITLE OF INVENTION: Lymphocyte Activation Antigen B-7 Family and
~ ; TITLE OF INVENTION: Polypeptides Encoded Thereby
~ ; FILE REFERENCE: 15966-562 (CURA-62)
~ ; CURRENT APPLICATION NUMBER: US/09/651,200
~ ; CURRENT FILING DATE: 2000-08-30
~ ; PRIOR APPLICATION NUMBER: 60/152383
~ ; PRIOR FILING DATE: 1999-09-03
~ ; PRIOR APPLICATION NUMBER: 60/172909
~ ; PRIOR FILING DATE: 1999-12-21
~ ; PRIOR APPLICATION NUMBER: 60/183578
~ ; PRIOR FILING DATE: 2000-02-18
~ ; NUMBER OF SEQUENCES: 25
~ ; SOFTWARE: Patentin Ver. 2.0

SEQ ID NO: 24
LENGTH: 534
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: Sequence
OTHER INFORMATION: mz5020.protein from Figure 4.
US-09-651-200-24

Query Match
Best Local Similarity   21.2%;   Score 123.5;   DB 4;   Length 534;
Matches   77;   Conservative   50;   Mismatches   128;   Indels   109;   Gaps   17;

```
Qy   14 SGSGNEVIEGPQNATV-LKGSQARFNCTVS--QGWKL---IMWALSDMVVLSVRPMEPI  66
           : :  : :  :: :  ::   :  :             ::IMWALSDMVVLSVRPMEPI
Db  242 SPTGAVEVQVPEDDVVALVGTDATLRCSFSPEPGFSLAQLNLIWQLTDTKQLV------- 294

Qy   67 ITNDRFTSQRYDQGGNF-------TSEMIHNVEPSDSGNIRCSLQNSRLHGS       112
           : ::                   :TSEMIHNVEPSDSGNIRCSLQNSRLHGS
Db  295 --HSFTEGR-DQGSAYANRTALFPDLLAQGNASLRLQRVRVADEGSFTCFV-SIRDFGS   349

Qy  113 AYLTVQVMGELFIPSVNLV----VAENEPCEVTCLPSHWTRLP--DISWELGL---LVSH 163
        ::  ::    ::                                 DISWELGL   LVSH
Db  350 AAVSLQVAAPYSKPSMTLEPNKDLRPGDTVTITC--SSYRGYPEAEVFWQDGQGVPLTGN 407

Qy  164 SSYYFVEPSDLQSAVSTLALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCP---QDT  220
           :                :::                           :   QDT
Db  408 VTTSQMANEQGLFDVHSVLRVVLGANGTYSC---------------LVRNPVLQQDA    449

Qy  221 GGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLIPTCTLTIRCCCCRRCCGNCCC  280
           : :                                                    CCC
Db  450 HGSVTIIGQPMFPPEAL-----WVTVGLSVCLIALLV---------------------- 482

Qy  281 RCCFCCCRRKRGFRIQFQKKSEKKTNKETETESGNENSGYNSDEQKTTETASLPPKSCES 340
           :  :                                                    CES
Db  483 ------------IKQSCEENAGAEDQDG------EGEGSKTALQPLKHSDS          525

Qy  341 SDPE 344
        : :
Db  526 KEDD 529
```

RESULT 12
US-08-633-148-4
Sequence 4, Application US/08633148
Patent No. 5864018
GENERAL INFORMATION:
APPLICANT: MORSER, MICHAEL J.
APPLICANT: NAGASHIMA, MARIKO
APPLICANT: HOLLANDER, DORIS A.
TITLE OF INVENTION: ANTIBODIES TO ADVANCED GLYCOSYLATION
TITLE OF INVENTION: END-PRODUCT RECEPTOR POLYPEPTIDES AND USES THEREFOR
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP
STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
CITY: SAN FRANCISCO

STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/08/633,148
FILING DATE: 16-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: MURPHY ESQ., MATTHEW B.
REGISTRATION NUMBER: 39,787
REFERENCE/DOCKET NUMBER: 014618-0056000US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
US-08-633-148-4

Query Match
Best Local Similarity   24.9%;   Score 122;   DB 2;   Length 318;
Matches   69;   Conservative   30;   Mismatches   86;   Indels   92;   Gaps   15;

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide

```
Qy    9 PEAVGGSGNEVIEGPQNATVLKGSQARFNCTVSQG--WKLIMWALSDMVVLSVRPMEP  65
         : :::                             : WKLIMWALSDMVVLSVRPMEP
Db  102 PEIVDSAS--ELTAGVPN------KVGTCVSEGSYPAGTLSWHLDG-------KP    141

Qy   66 IITNDRFTS------QRYDQGGNFT--SEMIHNVEPSDSGNIR----CSLQNSRLHGSAY 114
          :::::           :: :         :::  : :: ::        CSLQNSRLHGSAY
Db  142 LVPNEKGVSVKEQTRHIPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRAL 198

Qy  115 LTVQVMGELFIP----SVNLVVAENEP----------CEVTCLPSHWTRLPDISWE   156
        ::                    :           CEVTCLPSHWTRLPDISWE
Db  199 RTAPIQPRWEPVPLEEVQLVV--EPEGGAVAPGGTVTLTCEVPAQPS----PQIHWM  250

Qy  157 LGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRC 216
           :   :::                      :::CVATWKSLKARKSATVNLTVIRC
Db  251 KD-------GVPLPLPPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIE-  300

Qy  217 PQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGT 253
          : :
Db  301 PGEEG----------PTRAGSVGGSGLGT          318
```

RESULT 13
US-08-633-148-2
Sequence 2, Application US/08633148
Patent No. 5864018
GENERAL INFORMATION:
APPLICANT: MORSER, MICHAEL J.
APPLICANT: NAGASHIMA, MARIKO
APPLICANT: HOLLANDER, DORIS A.
TITLE OF INVENTION: ANTIBODIES TO ADVANCED GLYCOSYLATION
TITLE OF INVENTION: END-PRODUCT RECEPTOR POLYPEPTIDES AND USES THEREFOR
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP
STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/633,148
  FILING DATE: 16-APR-1996
  CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
  NAME: MURPHY ESQ., MATTHEW B.
  REGISTRATION NUMBER: 39,787
  REFERENCE/DOCKET NUMBER: 014618-0056000US
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (415) 326-2400
  TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 2:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 340 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: peptide
US-08-633-148-2

Query Match    5.9%;    Score 122;    DB 2;    Length 340;
Best Local Similarity   24.9%;    Pred. No. 0.0055;
Matches   69;    Conservative  30;    Mismatches  86;    Indels  92;    Gaps  15;

Qy    9  PEAVGSGSGNEVIEGPQNATVLKGSQARFNCTVSQG--WKLIMWALSDMVVLSVRPMEP    65
Db  124  PEIVDSAS--ELTAGVPN----------KVGTCVSEGSYPAGILSWHLDG-------KP   163

Qy   66  IIINDRFTS-----QRYDQGGNTT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGSAY  114
Db  164  LVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRAL  220

Qy  115  LTVQVMGELFIP------SVNLVVAENEP-------CEVTCLPSHWTRLPDISWE       156
Db  221  RTAPIQPRRWEPVPLEEVQLVV---EPEGGAVAPGGTVTLTCEVPAQPS----PQIHWM  272

Qy  157  LGLLVSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRC  216
Db  273  KD------------GVPLPPSPVLLIPEIGPDQOGTYSCVATHSSHGFQESRAVSISIIE-  322

Qy  217  PQDPGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGT  253
Db  323  PGEEG------------------------PTAGSVGGSGLGT  340

RESULT 14
US-08-374-834-1
Sequence 1, Application US/08374834
Patent No. 5656473
GENERAL INFORMATION:
  APPLICANT: Valenzuela, et al.
  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTOR
  NUMBER OF SEQUENCES: 17
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Regeneron Pharmaceuticals, Inc.
    STREET: 777 Old Saw Mill River Road
    CITY: Tarrytown
    STATE: New York
    COUNTRY: USA
    ZIP: 10591
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/374,834
    FILING DATE: 19-JAN-1995
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
```

```
  APPLICATION NUMBER: US 08/095,658
  FILING DATE: 21-JUL-1993
ATTORNEY/AGENT INFORMATION:
  NAME: Cobert, Robert J.
  REGISTRATION NUMBER: 36,108
  REFERENCE/DOCKET NUMBER: REG 190A
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (914) 345-7400
  TELEFAX: (914) 345-7721
INFORMATION FOR SEQ ID NO: 1:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 868 amino acids
  TYPE: amino acid
  STRANDEDNESS:
  TOPOLOGY: unknown
  MOLECULE TYPE: protein
US-08-374-834-1

Query Match    5.5%;    Score 114;    DB 1;    Length 868;
Best Local Similarity   21.3%;    Pred. No. 0.1;
Matches   80;    Conservative  40;    Mismatches 150;    Indels 106;    Gaps  18;

Qy   19  EVIEGPQNATVLKGSQARFNCTVSQGHK-LIMWALSDMVVLSVRPMEPIIITNDRFTSQRY  77
Db  122  KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGDSALRE---------NSRIAVLE-  171

Qy   78  DQGGNFISEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LITVQVMGELFIPSVNLVVAENE  136
Db  172  --------SGSLRIHNVQKEDAGQYRCVAKNSL-GTAYSKLVKLEVEVFARILRAPESHNV  223

Qy  137  P--------CEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQ  187
Db  224  TFGSFVTLRCTAIGMP------VPTISWIENGNAVSSGGSIQENVKDRVIDSRLQLFITKP-  277

Qy  188  SNGTLTCVAT----WKSLKARKSATVNLTVIR-------CPQDTGGGINIPGVLSSLP  234
Db  278  -GLYTCIATNKHGEKFSTAKAAATVSIAEWSKSQKESKGYCAQYRGEVCDAVLVKDSLV  335

Qy  235  SLGFSLP---------------TWGKVGLGLAGTMLLTPTCTLIIRCCCCRRRCCGCN----  277
Db  336  FFNTSYPDPEEAQELLIHTAWNEL-----KAVSPLCRPAAEAILCNHLFQECSPGVLP  388

Qy  278  P--------CCCRCC-----FCCRRKRGFRIQFQKKSEKEKTNKEFETESGNENSGYNSDBQKT  327
Db  389  TPMPICREYCLAVKELFCA-------------------KEWLAMEGKTHRGLIYRSGGMHF--LPV  431

Qy  328  TETASLP------PKSC  338
Db  432  PECSKLPSMHQDPTAC  447

RESULT 15
US-08-644-271-1
Sequence 1, Application US/08644271
Patent No. 5814478
GENERAL INFORMATION:
  APPLICANT: Valenzuela, et al.
  TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS
  TITLE OF INVENTION: AND LIGANDS
  NUMBER OF SEQUENCES: 32
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Regeneron Pharmaceuticals, Inc.
    STREET: 777 Old Saw Mill Road
    CITY: Tarrytown
    STATE: NY
    COUNTRY: USA
    ZIP: 10591
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Diskette
    COMPUTER: IBM Compatible
    OPERATING SYSTEM: DOS
    SOFTWARE: FastSEQ Version 2.0
  CURRENT APPLICATION DATA:
```

```
;
;  APPLICATION NUMBER:  US/08/644,271
;    FILING DATE:  10-MAY-1996
;    CLASSIFICATION:  435
;  PRIOR APPLICATION DATA:
;    APPLICATION NUMBER:  USSN 60/008,657
;    FILING DATE:  15-DEC-1995
;  ATTORNEY/AGENT INFORMATION:
;    NAME:  Cobert, Robert J
;    REGISTRATION NUMBER:  36,108
;    REFERENCE/DOCKET NUMBER:  REG 195A
;  TELECOMMUNICATION INFORMATION:
;    TELEPHONE:  914-345-7400
;    TELEFAX:  914-345-7721
;    TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;    LENGTH:  868 amino acids
;    TYPE:  amino acid
;    STRANDEDNESS:
;    TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-08-644-271-1

Query Match    5.5%;  Score 114;  DB 2;  Length 868;
Best Local Similarity  21.3%;  Pred. No. 0.1;
Matches  80;  Conservative 40;  Mismatches 150;  Indels 106;  Gaps 18;

Qy    19  EVIEGPQNATVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRPMEPIITNDRFTSQRY   77
          : :|| | ||| |:|| | | :  ||:: :
Db   122  KITRPPINVKIIEGLKAVLPCTTMGNDPKPSVSWIKGDSALRE----------NSRIAVLE-  171

Qy    78  DQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMGELFIPSVNLVVAENE  136
          :  | |||:  |||  |: || : |: | :|:| | |: |
Db   172  ------SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKLVKLEVVFARILRAPESHNV  223

Qy   137  P-----------CEVTCLPSHWTRLPDISW-ELGLLVSHSSYYFVPEPSDLQSAVSILALTPQ  187
              :|   ||  | |||:  | : || :
Db   224  TFGSFVVLRCTAIGMP-----VPTISWIENGNAVSSGSIQENVKDRVIDSRLQLFITKP-  277

Qy   188  SNGTLTCVAT----WKSLKARKSATVNLTVIR----------CPQDTGGGINIPGVLSSLP  234
          : |: | | :|:: :
Db   278  --GLYTCIATNKHGEKFSTAKAAATVSIAEWSKSQKESKGYCAQYRGEVCDAVLVKDSLV  335

Qy   235  SLGFSLP-----------------TWGKVGLGLAGTMLLTPTCTLYIRCCCCRRRCCGCN-----  277
                               :|  ||  |  :
Db   336  FENTSYPDPEEAQELLIHTAWNEL------KAVSPLCRPAAEALLCNHLFQECSPGVLP  388

Qy   278  ----CCCRCC--------FCCRRKRGFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKT  327
             | |  ||: :| : | |: ||: :||:::
Db   389  TPMPICRBYCLAVKELFCA----------------KEWLAMEGKTHRGLYRSGMHF--LPV  431

Qy   328  TETASLP--------PKSC  338
              : :|| : :||
Db   432  PECSKLPSMHQDPTAC  447

Search completed: April 28, 2003, 21:12:23
Job time : 18.0485 secs
```

OM protein - protein search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 16:24:31 ; Search time 38.4662 Seconds
(without alignments)
1337.141 Million cell updates/sec

Title: US-09-729-264-6
Perfect score: 2077
Sequence: 1 MERHLLTVPEAVSGSGNEV...........HPQASFNLASPEKVSNTTVV 386
Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database : A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2077	100.0	386	23	AAU75542	Human B7-like prot
2	2014	97.0	382	23	AAU75540	Human B7-like prot
3	2000	96.3	386	23	AAU75541	Human B7-like prot
4	1514	72.9	377	23	AAU75543	Human B7-like prot
5	1326	63.8	463	22	ABG28169	Novel human diagno
6	928.5	44.7	370	23	AAU75544	Mouse B7-like prot
7	587.5	28.3	631	23	AAU75547	Rat B7-like protei
8	579.5	27.9	270	23	AAU75545	Mouse B7-like prot
9	498	24.0	223	23	AAU75546	Mouse B7-like prot
10	151	7.3	404	22	AAB81925	Extracorporeal cir
11	151	7.3	404	23	AAE23219	Human receptor for
12	151	7.3	404	23	AAU77543	Human receptor for
13	151	7.3	404	23	AAM48745	Human RAGE protein
14	151	7.3	404	20	AAW81030	Melanoma associate
15	151	7.3	1496	21	AAY70469	Human p53 target m
16	135	6.5	1498	22	ABB11587	Human peroxidasin
17	134	6.5	592	23	ABB75751	Human immunoglobul
18	134	6.5	594	23	ABB75753	Human pancreas GP3
19	134	6.5	708	22	AAU29315	Human PRO polypept
20	134	6.5	708	23	ABB75752	Human gp354 (putat
21	134	6.4	344	22	AAM38713	Human polypeptide
22	133.5	6.2	1447	16	AAR68953	Deleted in colorec
23	128	6.2	1447	20	AAY33498	Deleted in colorec
24	128	6.2	1953	23	AAR13144	Human UNC-40 prote
25	128	6.2	467	12	AAU08351	Deleted in Colorec
26	128	6.2	475	23	AAW62575	Protein MYLK diffe
27	128	6.2	863	17	AAR94982	Alternatively spli
28	128	6.1	863	19	AAW62569	Alternatively spli
29	127.5	6.1	869	18	AAW62583	Nsk2 extracellular
30	127.5	6.1	869	19	AAW26611	Alternatively spli
31	127.5	6.1	869	18	AAW25506	Mouse receptor tyr
32	127.5	6.1	871	18	AAR84087	Human muscle-speci
33	127.5	6.1	871	17	AAW62568	Human Dmk receptor
34	127.5	6.1	871	17	AAR84092	Nsk2 receptor. Mu
35	127.5	6.1	873	19	AAW62573	Mouse receptor tyr
36	127.5	6.1	873	19	AAR84091	Nsk2 receptor with
37	127.5	6.1	881	17	AAW62572	Alternatively spli
38	127.5	6.1	881	19	AAY52130	Nsk2 receptor with
39	127	6.1	332	21	AAG77857	Mouse Nsk2 (altern
40	127	6.0	537	22	AAG77858	Mutant protein mMu
41	127	6.0	576	22	AAG77856	Mutant protein mMu
42	126	6.0	869	22	AAB68421	Mutant protein of
43	125.5	6.0	869	22	AAW08747	Protein of muscle
44	125.5	6.0	1070	18	ABG61942	Amino acid sequenc
45	125	6.0	1070	23		Human colon carcin

ALIGNMENTS

RESULT 1
AAU75542

ID   AAU75542 standard; Protein; 386 AA.
AC   AAU75542;
XX
DT   23-APR-2002 (first entry)
XX
DE   Human B7-like protein, B7-L_h3.
XX
OS   Homo sapiens.
XX
PN   WO20020710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.

XX
DR WPI; 2002-130881/17.
DR N-PSDB; ABKI3030.
XX
PT New B7-like polypeptides, polynucleotides and their modulators, useful
PT for diagnosing, preventing and treating reproductive, immune and
PT proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS Claim 13; Fig 3; 135pp; English.
XX

CC The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC The polypeptide, polynucleotide encoding it and antibody against (I) are
CC useful for treating B7-like polypeptide-related disease, disorders or
CC conditions including reproductive disorders (e.g. infertility,
CC miscarriage, preterm labour and delivery and endometriosis) and
CC proliferative disorders. Antibodies, soluble proteins comprising
CC extracellular domains and other regulators of B7-L polypeptides are
CC useful for enhancing the immune response to tumours. (I) plays a role in
CC growth and maintenance of cancer cells based on the observation of
CC seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC polypeptide. Hence modulators of (I) are useful for the treatment of
CC cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC in allograft transplantation, graft versus host disease, T-cell
CC dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC molecules are useful for alleviating the symptoms associated with
CC diseases involving chronic immune cell dysfunction or to treat
CC autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC purpura and psoriasis, chronic inflammatory disease such as
CC inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC are also useful as immunosuppressive agents for bone marrow and organ
CC transplantation or to prolong graft survival. B7-L molecules are also
CC useful for diagnosis and treatment of diseases involving abnormal cell
CC proliferation, including arteriosclerosis and vascular restenosis.
CC Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC shock syndrome or allosensitisation due to blood transfusions, and for
CC treatment of allergy, asthma and hypersensitivity reactions, various
CC nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC lymphoproliferative disorders such as multiple myeloma. The present
CC sequence represents the amino acid sequence of human B7-L_h3.
XX
SQ Sequence 386 AA;

Query Match   100.0%; Score 2077; DB 23; Length 386;
Best Local Similarity   100.0%;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy   1 MERHLLTVPEAVGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSV 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MERHLLTVPEAVGSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSV 60

Qy   61 RPMEPIITNDRFTSQRYDQGGNTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVM 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 RPMEPIITNDRFTSQRYDQGGNTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVM 120

Qy   121 GELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVS 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYFVPEPSDLQSAVS 180

Qy   181 ILALTPQSNGTLTCVATWKSLRARKSATVNLTIVIRCPQDTGGGINIPGVLSSLPSLGFSL 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ILALTPQSNGTLTCVATWKSLRARKSATVNLTIVIRCPQDTGGGINIPGVLSSLPSLGFSL 240

Qy   241 PTWGKVGLGLAGTMLLTPTCTLFIIRCCCCRRCCGCNCCCRCCFCCRRKRGFRIQFQKKS 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 PTWGKVGLGLAGTMLLTPTCTLFIIRCCCCRRCCGCNCCCRCCFCCRRKRGFRIQFQKKS 300

Qy   301 EKEKTNKEETETESGNENSGYNSDEOKTTETFASLLPPKCSCESSDPEQRNSSCGPPHQRADQR 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 EKEKTNKEETETESGNENSGYNSDEOKTTETFASLLPPKCSCESSDPEQRNSSCGPPHQRADQR 360

Qy   361 PPRPASHPQASFNLASPEKVSNTTVV 386
          ||||||||||||||||||||||||||
Db   361 PPRPASHPQASFNLASPEKVSNTTVV 386

RESULT 2
AAU75540
ID AAU75540 standard; Protein; 382 AA.
XX
AC AAU75540;
XX
DT 23-APR-2002 (first entry)
XX
DE Human B7-like protein, B7-L_h1.
XX
KW Human; B7-like protein; B7-L; antinfertility; gynaecological;
KW antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW reproductive disorder; graft versus host disease; autoimmune disease;
KW toxic shock syndrome; allergy; nephropathy; skin disorder;
KW endocrinopathy; lymphoproliferative disorder.
XX
OS Homo sapiens.
XX
PN WO200200710-A2.
XX
PD 03-JAN-2002.
XX
PF 28-JUN-2001; 2001WO-US20719.
XX
PR 28-JUN-2000; 2000US-214512P.
PR 28-NOV-2000; 2000US-0729264.
XX
PA (AMGE-) AMGEN INC.
XX
PI Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR WPI; 2002-130881/17.
DR N-PSDB; ABKI3028.
XX
PT New B7-like polypeptides, polynucleotides and their modulators, useful
PT for diagnosing, preventing and treating reproductive, immune and
PT proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS Claim 13; Fig 1; 135pp; English.
XX

CC The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC The polypeptide, polynucleotide encoding it and antibody against (I) are
CC useful for treating B7-like polypeptide-related disease, disorders or
CC conditions including reproductive disorders (e.g. infertility,
CC miscarriage, preterm labour and delivery and endometriosis) and
CC proliferative disorders. Antibodies, soluble proteins comprising
CC extracellular domains and other regulators of B7-L polypeptides are
CC useful for enhancing the immune response to tumours. (I) plays a role in
CC growth and maintenance of cancer cells based on the observation of
CC seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC polypeptide. Hence modulators of (I) are useful for the treatment of
CC cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC in allograft transplantation, graft versus host disease, T-cell
CC dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC molecules are useful for alleviating the symptoms associated with
CC diseases involving chronic immune cell dysfunction or to treat
CC autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC purpura and psoriasis, chronic inflammatory disease such as

```
CC   Inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of human B7-L_h1.
XX
SQ   Sequence 382 AA;

Query Match       Best Local Similarity   97.0%;
                  Conservative   99.9%;   Score 2014;  DB 23;  Length 382;
Matches  373;                    1;  Pred. No. 3.3e-159;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   13 GSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9 GSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 68

Qy   73 TSQRYDQGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV 132
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69 TSQRYDQGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV 128

Qy  133 AENEPCEVTCLPSHWTRLPDISWELGLLVSHSSSYFVPEPSDLQSAVSILALTPQSNGTL 192
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  129 AENEPCEVTCLPSHWTRLPDISWELGLLVSHSSSYFVPEPSDLQSAVSILALTPQSNGTL 188

Qy  193 TCVATWKSLRARKSATVNLTVIRCPDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG 252
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  189 TCVATWKSLRARKSATVNLTVIRCPDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG 248

Qy  253 TMLLIPTCLLTIRCCCCRRCCGNCNCCCRRCCGFRIQFOKKSEKEKTNKETETE 312
         ||||||||||||||||         ||||||||||||||||||||||||||||
Db  249 TMLLIPTCLLTIRCCCRRCCGNCNCCCRRCCFCCRRKRGFRIQFOKKSEKEKTNKETETE 308

Qy  313 SGNENSGYNSDEQKTETASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASF 372
         |||||||||||| ||| |||||||||||||||||||||||||||||||||||||||||
Db  309 SGNENSGYNSDEQKTTDTASLPPKSCESSDPEQRNSSCGPPHQRADQRPPRPASHPQASF 368

Qy  373 NLASPEKVSNTTVV 386
         |||||||||||||
Db  369 NLASPEKVSNTTVV 382
```

```
RESULT 3
AAU75541
ID   AAU75541 standard; protein; 386 AA.
XX
AC   AAU75541;
XX
DT   23-APR-2002  (first entry)
XX
DE   Human B7-like protein, B7-L_h2.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; antineuroprotective;
KW   antidiabetic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Homo sapiens.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis --
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13029.
XX
PS   Claim 13; Fig 2; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I), are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of human B7-L_h2.
XX
SQ   Sequence 386 AA;

Query Match       Best Local Similarity   96.3%;
                  Conservative   99.2%;   Score 2000;  DB 23;  Length 386;
Matches  371;                    1;  Pred. No. 4.8e-158;  Mismatches   2;  Indels   0;  Gaps   0;

Qy   13 GSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72
         |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db   13 GSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72

Qy   73 TSQRYDQGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV 132
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   73 TSQRYDQGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV 132

Qy  133 AENEPCEVTCLPSHWTRLPDISWELGLLVSHSSSYFVPEPSDLQSAVSILALTPQSNGTL 192
```

```
         |||||||||||
         |||||||||||
Db  133  AENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTL 192
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  193  TCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG 252
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193  TCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG 252
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  253  TMLLTPTCTLTIRCCCCRRRCCGCNCCCRRCCFCCRRKRGFRIQFQKKSEKEKTNKETETE 312
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  253  TMLLTPTCTLTIRCCCCRRRCCGCNCCCRRCCFCCRRKRGFRIQFQKKSEKEKTNKETETE 312
         |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Qy  313  SGNENSGYNSDEQKTFETASLPPKSCESSSDPEQRNSSCGPHQRADQRPRPRASHPQASF 372
Db  313  SGNENSGYNSDEQKTDTIASLPPKSCESSSDPEQRNSSCGPHQRADQRPPRPASHPQASF 372
         |||||||||||||
Qy  373  NLASPEKVSNTTVV 386
         |||||||||||||
Db  373  NLASPEKVSNTTVV 386
```

RESULT 4
AAU75543
ID   AAU75543 standard; protein; 377 AA.
XX
AC   AAU75543;
XX
DT   23-APR-2002 (first entry)
XX
DE   Human B7-like protein, B7-L_h4.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
OS   Homo sapiens.
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13031.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
PS   Claim 13; Fig 4; 135pp; English.
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L

CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of human B7-L_h4.
XX
SQ   Sequence 377 AA;

Query Match   72.9%; Score 1514; DB 23; Length 377;
Best Local Similarity   88.2%; Pred. No. 1.3e-117;
Matches 284; Conservative 12; Mismatches 20; Indels 6; Gaps 2;

```
Qy   13  GSGSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72
         |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   13  GSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72
Qy   73  TSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV 132
         ||||||||||  |||||||||||||||:||||||||||||||||||||||||||||||||
Db   73  TSQRYDQGGNLTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLVV 132
Qy  133  AENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTL 192
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  133  AENEPCEVTCLPSHWTWLPDISWELGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTL 192
Qy  193  TCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG 252
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  193  TCVATWKSLKARKSATVNLTVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAG 252
Qy  253  TMLLTPTCTLTIRCCCCRRRCCGCNCCCRRCCFCCRRKRGFRIQFQKKSEKEKTNKETETE 312
         |||||||||||||||||||||||||||||||||||||||:|:|
Db  253  TMLLTPTCTLTIRCCCCRRRCCGCNCCCRRCCFCCRRKRG---NLKKRRQTRKLRQKVEMK 309
Qy  313  SGNENSGYNSDEQKTFETASLP 334
         :::  |||               |||
Db  310  T---PATIQMNKRPQTPLLSLP 328
```

RESULT 5
ABG28169
ID   ABG28169 standard; protein; 463 AA.
XX
AC   ABG28169;
XX
DT   18-FEB-2002 (first entry)
XX
DE   Novel human diagnostic protein #28160.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
OS   Homo sapiens.
XX
XX

PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS92356.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID NO 58528; 103pp; English.

CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II), (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   463 AA;

RESULT 6

Query Match   63.8%; Score 1326; DB 22; Length 463;
Best Local Similarity   100.0%; Pred. No. 7.3e-102;
Matches 243; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   49 MWALSDMVVLSVREMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSR 108
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MWALSDMVVLSVREMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSR 60

Qy   109 LHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYF 168
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 LHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSYYF 120

Qy   169 VPEPSDLQSAVSILALTPQSNGTLTCVATWKSLRARKSATVNLTVIRCPQDTGGGINIPG 228
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 VPEPSDLQSAVSILALTPQSNGTLTCVAATWKSLRARKSATVNLTVIRCPQDTGGGINIPG 180

Qy   229 VLSSLPSLGFSLPTWGKVGLGLAGTMLLIPPTCTLTIRCCCCRRCCGCNCCCRCCFCCRR 288
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 VLSSLPSLGFSLPTWGKVGLGLAGTMLLIPPTCTLTIRCCCCRRCCGCNCCCRCCFCCRR 240

Qy   289 KRG 291
              |||
Db   241 KRG 243

ID   AAU75544 standard; protein; 370 AA.
XX
AC   AAU75544;
XX
DT   23-APR-2002 (first entry)
XX
DE   Mouse B7-like protein, B7-L_m1.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Mus musculus.
XX
PN   WO20020710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13032.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 13; Fig 5; 135pp; English.

CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions, and for
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and

CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of mouse B7-L_m1.
XX
SQ   Sequence   370 AA;

Query Match   44.7%; Score 928.5; DB 23; Length 370;
Best Local Similarity   50.0%; Pred. No. 6.4e-69;
Matches 194; Conservative 54; Mismatches 105; Indels 35; Gaps 5;

Qy   5 LLTVPEAVGSGSGGNEVIEGPQNATVLKGSQARPNCTVSQGWKLIMWALSDMVVLSVRPME 64
            ||| ||||||||||| |:| || :| ||||||:||:|| |:||||:
Db   12 LVILAQLTASGGSSYQIIEGPQNYTVLKDSEAHFNCTVTHGWKLLMWTLNQMVVLSTTQG 71

Qy   65 PIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLIYQVMGELF 124
            ||| |:   ::|||||||| |||||||||||||:|||||||||:|||:
Db   72 PIITNNRRTYASVNSTDSFISELIIHDVQPSDSGSVQCSLQNSHGFGSAFLSVQVMGTLN 131

Qy   125 IPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSSYYFVEPSDLQSAVSIIAL 184
            ||| |:|| | |||||||||:|||||||||||:| || |:||||
Db   132 IPSNNLIVTEGEPCNVTCYAVGWTSLPDISWELEVPVSHSSYNSFLEPGNFMRVLSVDL 191

Qy   185 TPQSNGILTCVAIMKSLKARKSATVNLTVIRCPQDTGGINIPGVLSSLPSLGFSLPTWG 244
            || | ||| | || |:||||||||||| :||||||| |||| |||
Db   192 TPLGNGTLTCVAELKDLQASKLTVNLTVVQPPPD---------------SIGEEGPALPTWA 239

Qy   245 KVGLGLAGTMLLTPTCLTLTRCCCCRRRCCGCNCCCRCCFCCCRRKRGFRIQFQ---KKSE 301
            :|||||| |:| |:||||||||||||||| ||||||| |||||:| |||
Db   240 IILLAVAFGLLLILIVLIIFCCC------------CASRREKEESTYQNEIRKSA 284

Qy   302 KEKTNK---ETETESGNENSGNYNSDEQKTTETASLPPKSCESSDPEQRNSSCGPPHQRAD 358
            :||| |||||||:|||||||||||||||||||||||| |||: ||:
Db   285 NMRTNKADPETKLKSGKENYGYSSDEAKAAQTASLPPKSAEVSLPERKSSSSL--PYQELN 342

Qy   359 QRPRPRPASHPQASFNLASPEKYVSNTTVV 386
            : |||: ||: ||:||||:||:||:|
Db   343 KHQPGPATHPRVSFDIASPQKVRNVTLV 370

RESULT 7
AAU75547
ID   AAU75547 standard; Protein; 631 AA.
XX
AC   AAU75547;
XX
DT   23-APR-2002 (first entry)
XX
DE   Rat B7-like protein, B7-1.
XX
PS   Disclosure; Fig 8; 135pp; English.
XX
PN   WO0200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX

PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility),
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   in allograft transplantation, graft versus host disease, T-cell
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions.
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of rat B7-L.
XX
SQ   Sequence   631 AA;

Query Match   28.3%; Score 587.5; DB 23; Length 631;
Best Local Similarity   26.4%; Pred. No. 3e-40;
Matches 170; Conservative 47; Mismatches 112; Indels 315; Gaps 10;

Qy   5 LLTVPEAVGSGSGGNEVIEGPQ------------------------------------ 25
            ||| |||||||||||:
Db   12 LVILAQLTASGGSSYQIIEGPQMAYSCQPLQESPLLGFPRLRFIHLFVLLLQISSGI 71

Qy   26 --------------NATVLKGSQARNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRF 72
                          | ||||:| |||||||||||||||:|| |:||||:|||||||
Db   72 VGQVSKSVRNVTVLKDSEAHFNCTVTHGWKLLMWTLNQMVVLSTTQGPIITNRFEKAL 131

Qy   73 -----------------------------TSQRYDQGGNFTSEMI 88
                                         ||||||||||||||:|
Db   132 LSCDYKFCSEEQSIHRIYWQKHDKMVLSVISGVPEVWPKYKNRTIYASVNSTDSFISELI 191

Qy   89 IHNVEPSDSGNIRCSLQNSRLHGSAYL---------- 115
            |:|:||||||||||||||||||:||||
Db   192 IHDVQPSDSGSVQCSLQNSHGFGSAFLSVQVDIANNYSFSLLGLLLSDRGTYTCVQRY 251

Qy   116 -------------TVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHS 164
                         |||:||||| |||||||||||||||||||||||||||||||||||
Db   252 EGGSYVVKHLITVEVMGTLNIPSNNLIVTEGEPCNVTCYAVGWTSLPDISWELEVPVSHS 311

Qy   165 --SYYFPEPSDLQ 176
                                                                         || :
Db   312 LSVRADFPTPNTIEYGNPSADIKRITCFASGGPPKPRLSWLENGRELNSYNSFLEPGNFM 371

Qy   177 SAVSTLALPPQSNGTLTCVATWKSLKARKSATVNL------------------------- 211
           :|:| ||||||||||||||||||||||
Db   372 RVLSVLDIPLGNGTLTCVAELKLDLQASKSLTVNLGINTTISQDPESELYTISSQLDFNA 431

Qy   212 -------------TVIRCPQDTGGGINIPGVLSSLPFSLGFSLPTWGKVGLGL 250
                        |:|:| ||:| | | ::|||| ||| :|
Db   432 TYDHFIDCFIEYGDAHVSQNFTVVQPPPD-----------SIGEEGPALPTWAIILLAV 479

Qy   251 AGTMLL--TPT 259
             |:||
Db   480 AFSLLLLILIVLLIIFTWVKPPEDPPDEKQTVPFAWAGPDAVKAIIIFFIAIVIAVIAA 539

Qy   260 CTLITIRCCCCRRRCCGCNCCCRRGFRIQFQ-KKSEKETKNK---ETETESGN 315
             :|| ||||||||| ||| |||||| :|||
Db   540 IAIIIFCCCCASR------------REKEESTYQNEIRKSANMRTNKADPETKLKSGK 585

Qy   316 ENSGYNSDEQKTTETASLPPKSCESSDPEQRNSSCGPPHQRADQ 359
             ||||||||||
Db   586 ENYGYSSDEQCITVKFRRCFRRRNEASREINKNLYIGPVEAAAEQ 629

RESULT 8
AAU75545
ID   AAU75545 standard; Protein; 270 AA.
XX
AC   AAU75545;
XX
DT   23-APR-2002 (first entry)
XX
DE   Mouse B7-like protein, B7-L_m2.
XX
OS   Mus musculus.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
CC   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory bowel disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the amino acid sequence of mouse B7-L_m2.
XX
SQ   Sequence   270 AA;

Query Match   27.9%; Score 579.5; DB 23; Length 270;
Best Local Similarity   34.3%; Pred. NO. 4.7e-40;
Matches 113; Conservative 46; Mismatches 74; Indels 135; Gaps 5;

Qy   5 LLIVPEAVGSGSGNEVIEGPQNATIVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPME 64
           |: : | :||||||| |: ||||||||:|||||:|||||:|:|:| :|||||
Db   12 LVILAQLITASGSSYQIIEGPQNVTVLKDSEAHPNCTVTHGWKLLMWLNQMVVLSITTQG 71

Qy   65 PIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQVMGELF 124
           ||||||||||| | :|| |||| |:|||||:|||||:|||||||||||:|||||||
Db   72 PIITNNRFIYASYNSTDSFPISELLIHDVQPSDSGSVQCSLQNSHGFGSAFLSVQ----- 125

Qy   125 IPSVNLVAENEPCEVTCLPSHWTRLPDISWELGLLVSHSSSYYFVPFEPSDLQSAVSIAL 184
           :|||||||||||||||||||:||||||||:||||||||:|:||||||||||:|||||||
Db   126 --CASRREKEESTYQNEIRKSA 184

Qy   185 TPQSNGTLTCVATWKSLKARKSATVNLTIVIRCPQDTGGGINIPGVLSSLPSLGFSLPTWG 244
           ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db   185 NMRTNKADPETKLKSGKENYGYSSDDEAKAAQTASLPKSAEVSLPERKSSSL--PYQELN 242

Qy   245 KVGLGLAGTMLLTPTCTLITIRCCCCRRRCCGCNCCCRRGFRIQFQ---KKSE 301
           |||||| | | :||| || |||||||||||||||||||||||| ||||
Db   243 KHQPGPATHPRVSPDIASPQKVRNVTLV 270

RESULT 9
AAU75546
ID   AAU75546 standard; Protein; 223 AA.
XX
AC   AAU75546;

XX
PA   (AMGE-) AMGEN INC.
PA   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
PI
PN   WO20020710-A2.
PD   03-JAN-2002.
PF   28-JUN-2001; 2001WO-US20719.
DR   WPI; 2002-130881/17.
DR   N-PSDB; ABK13033.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PS   Claim 13; Fig 6; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and

```
DT   23-APR-2002 (first entry)
XX
DE   Mouse B7-like protein, B7-L_m3.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; antipsoriatic; dermatological; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder.
XX
OS   Mus musculus.
XX
PN   WO20020071O-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI: 2002-130881/17.
DR   N-PSDB; ABK13034.
XX
PS   Claim 13; Fig 7; 135pp; English.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
```

CC The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC The polypeptide, polynucleotide encoding it and antibody against (I) are
CC useful for treating B7-like polypeptide-related disease, disorders or
CC conditions including reproductive disorders (e.g. infertility,
CC miscarriage, preterm labour and delivery and endometriosis) and
CC proliferative disorders. Antibodies, soluble proteins comprising
CC extracellular domains and other regulators, of B7-L polypeptides are
CC useful for enhancing the immune response to tumours. (I) plays a role in
CC growth and maintenance of cancer cells based on the observation of
CC seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC polypeptide. Hence modulators of (I) are useful for the treatment of
CC cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC in allograft transplantation, graft versus host disease. T-cell
CC dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC molecules are useful for alleviating the symptoms associated with
CC diseases involving chronic immune cell dysfunction or to treat
CC autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC purpura and psoriasis, chronic inflammatory disease such as
CC inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC are also useful as immunosuppressive agents for bone marrow and organ
CC transplantation or to prolong graft survival. B7-L molecules are also
CC useful for diagnosis and treatment of diseases involving abnormal cell
CC proliferation, including arteriosclerosis and vascular restenosis.
CC Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC shock syndrome or allosensitisation due to blood transfusions, and for
CC treatment of allergy, asthma and hypersensitivity reactions,
CC nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC pemphigoid) endocrinopathies (Grave's disease), various pneumopathies
CC (extrinsic alveolitis), vasculopathies (Grave's disease, anaemia,
CC thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC lymphoproliferative disorders such as multiple myeloma. The present
CC sequence represents the amino acid sequence of mouse B7-L_m3.

```
XX
SQ   Sequence 223 AA;
XX

                 Query Match    24.0%;  Score 498;  DB 23;  Length 223;
                 Best Local Similarity   31.2%;  Pred. No. 2.2e-33;
                 Matches 120;  Conservative 41;  Mismatches 48;  Indels 176;  Gaps 6;

Qy     5 LLTVPEAVGSGSGNEVIEPQNATVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPME  64
           ||| :: ||||| : ||||||| ||||  ::||||||||| :: |||||  | ||||:::
Db    12 LVILAQLTASGSSYQIIEGPQNVTVKDSEAHFNCTVTHGWKLLMWTLNQMVVLSLTTQG  71

Qy    65 PIITNDRFTSQRYDQGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAXILTVQVMGELF 124
           |||| |||| : :: ||| ::|||| ||:||||| :|:||||||:| ||::|:|
Db    72 PIITNNRFTYASYNSTDSFISELIIHDVQPSDSGSGSVQCSLQNSHGFGSAFLSVQ---- 125

Qy   125 IPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLIVSHSSSYYFVPEPSDLQSAVSILAL 184
                                                  :|: ::
Db   126 -------------------------------------ESTY----QNEI----------- 133

Qy   185 TPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDFGGGINIPGVLSSLPSLGFSLPTWG 244
                                                            |||:
Db   134 --------------------------------------------------RKSANM---- 139

Qy   245 KVGLGLAGTMLLITPTCTLTIRCCCRRRCCGCNCCCRCCFCCRRKRGFRIIQFKKSEKEK 304
                                                                       R
Db   140 ----------------------------------------------------------R- 140

Qy   305 TNK--ETETESGNENSGYNSEDEQKTTETASLPPKSCCESSDPEQRNSSCGPPHQRADQRP 361
           |||   : : :||||:| ::|| ::|: ||||||||| |: |||: | : : ||
Db   141 TNKADPETKLKSGKENYGYSSDEAKRAAQTASLPPKSAEVSLPEKRSSSL--PYOELNKHQ 198

Qy   362 PRPASHPQASFNLASPEKVSNTTVV 386
           | ||:||: ||:|||||:|:| ::|
Db   199 PGPATHPRVSFDIASPQKVRNVTLV 223
```

RESULT 10

```
AAB81925
ID   AAB81925 standard; protein; 404 AA.
XX
AC   AAB81925;
XX
DT   15-JUN-2001 (first entry)
XX
DE   Extracorporeal circulation material receptor protein.
XX
KW   Extracorporeal circulation; carbonyl stress product; receptor;
KW   diabetes; vascular lesion; excretory dysfunction.
XX
OS   Unidentified.
XX
PN   WO200118060-A1.
XX
PD   15-MAR-2001.
XX
PF   08-SEP-2000; 2000WO-JP06172.
XX
PR   08-SEP-1999; 99JP-0254463.
XX
PA   (TORA ) TORAY IND INC.
XX
PI   Shimizu S, Kubota M, Akiyama H, Usui M;
XX
DR   WPI: 2001-290314/30.
XX
PT   Material for extracorporeal circulation, applicable in selective
PT   elimination of diabetic complication factors such as carbonyl stress
PT   products caused by abnormally promoted carbonyl stress from excretory
PT   dysfunction in vascular lesions -
XX
PS   Claim 1; Page 31-32; 36pp; Japanese.
XX
```

CC The present invention describes a material for extracorporeal circulation

CC which is made from a water;insoluble carrier immobilized with a protein
CC having the sequence shown here. The materials of the invention, including
CC adsorbents, are for extracorporeal circulation, which are applicable in
CC the selective elimination of diabetic complication factors from a body
CC fluid, and are therefore useful in treating vascular lesions like
CC arteriosclerosis due to carbonyl stress products caused by abnormally
CC promoted carbonyl stress from excretory dysfunction.
XX
SQ Sequence   404 AA;

Query Match   7.3%;   Score 151; DB 22; Length 404;
Best Local Similarity   23.3%; Pred. No. 0.00035;
Matches   90; Conservative 38; Mismatches 112; Indels 146; Gaps 19;

Qy   9 PEAVGSGSGNEVIEGFPQNATVLKGSQARFNCTVSQG---WKLIMWALSDMVVLSVRPMEP 65
           || :|: | |: | : :| | | :|
Db   124 PEIVDSAS--ELTAGVPN-------KVGTCVSEGSYPAGTLSWHLDG-------KP 163

Qy   66 IITNDRFTS-----QRYDQGGNFT--SEMIHNVEPSDSGNIR----CSLQNSRLHGSAY 114
           ::| ::|:   :|
Db   164 LVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRAL 220

Qy   115 LTVQVMGELFIP-----SVNLVVAENEP----------------CEVTCLPSHWTRLPDISWE 156
           :| : : |: :| :||
Db   221 RTAPIQPRVWEPVPLEEVQLVV---EPEGGAVAPGGTVTLICEVPAQPS-----PQIHWM 272

Qy   157 LGLLVSHSSSYYFVPEPSDLQSAVSILALTPQSNGTLITCVATWKSLKARKSATVNLTVIRC 216
           ||| :   :: | | | | |: :| : :| | :|:|
Db   273 KD----------GVPLPLPPSPVLILPEIGPQDOGTYSCVATHSSHGPQESRAVSISTIE- 322

Qy   217 PQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLT------PTCTLTIRCCCC 269
           | |||:||:| | |:||| :| ||| | | || :|:: | | ||:|||
Db   323 PGEEG--------------PTAGSVGGSGLGTLALALGILGGLGTAALLIGVILW 363

Qy   270 RRRCCGCNCCCRCCFCCRRKRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQKT 327
           :|: |: :||:||||
Db   364 QRR-----------------QRRGEERKAPENQEEEERAELN----------- 389

Qy   328 TETASLPPKSCESSSDPEQRNSSCGPP 353
           :| :|| |:||: |||
Db   390 ----------QSEEPEAGESSTGP 404

RESULT 11
AAE23219
ID   AAE23219 standard; Protein; 404 AA.
XX
AC   AAE23219;
XX
DT   27-AUG-2002 (first entry)
XX
DE   Human receptor for advanced glycosylation end product (RAGE) protein.
XX
KW   Human; Receptor for advanced glycosylation end product; RAGE; cardiant;
KW   tissue growth; neointimal formation; blood vessel; restenosis; diabetes;
KW   myocardial infarction; angioplasty; peripheral vascular surgery; angina;
KW   transgenic animal; acute thrombotic stroke; venous thrombosis.
XX
OS   Homo sapiens.
XX
XX
PN   WO200230889-A2.
XX
PD   18-APR-2002.
XX
PF   12-OCT-2001; 2001WO-US32036.
XX
PR   13-OCT-2000; 2000US-0687528.
XX
PA   (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI   Stern DM, Schmidt A, Marso S, Topol E, Lincoff AM;
XX
DR   WPI; 2002-426260/45.

DR   N-PSDB; AAD36952.
XX
     Disclosure; Page 16; 43pp; English.
XX
PT   Inhibiting new tissue growth or neointimal formation in blood vessels
PT   of subject suffering from diabetes, stroke and preventing restenosis,
PT   comprises administering inhibitor of receptor for advance glycation end
PT   product.
XX
CC   The invention relates to a method for inhibiting new tissue growth or
CC   neointimal formation in blood vessels in a subject that has experienced
CC   blood vessel injury and preventing exaggerated restenosis in a diabetic
CC   subject. The method comprises administering an inhibitor of receptor for
CC   advanced glycation/glycosylation end product (RAGE), so as to inhibit new
CC   tissue growth or neointimal formation in subject's blood vessels and
CC   preventing restenosis in the subject. The method is useful for inhibiting
CC   new tissue growth or neointimal formation in blood vessels in a subject
CC   like non-human animal, a transgenic non-human animal or a human suffering
CC   from diabetes, acute thrombotic stroke, venous thrombosis, unstable
CC   angina, myocardial infarction, abrupt closure following angioplasty or
CC   stent placement, or thrombosis as a result of peripheral vascular surgery
CC   The method is also useful for preventing restenosis and for determining
CC   whether a compound inhibits new tissue growth in a blood vessel in a
CC   subject. The present sequence is human receptor for advanced
CC   glycosylation end product (RAGE) protein.
XX
SQ   Sequence   404 AA;

Query Match   7.3%;   Score 151; DB 23; Length 404;
Best Local Similarity   23.3%; Pred. No. 0.00035;
Matches   90; Conservative 38; Mismatches 112; Indels 146; Gaps 19;

Qy   9 PEAVGSGSGNEVIEGFPQNATVLKGSQARFNCTVSQG---WKLIMWALSDMVVLSVRPMEP 65
           || :|: | |: | : :| | | :|
Db   124 PEIVDSAS--ELTAGVPN-------KVGTCVSEGSYPAGTLSWHLDG-------KP 163

Qy   66 IITNDRFTS-----QRYDQGGNFT--SEMIHNVEPSDSGNIR----CSLQNSRLHGSAY 114
           ::| ::|:   :|
Db   164 LVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRAL 220

Qy   115 LTVQVMGELFIP-----SVNLVVAENEP----------------CEVTCLPSHWTRLPDISWE 156
           :| : : |: :| :||
Db   221 RTAPIQPRVWEPVPLEEVQLVV---EPEGGAVAPGGTVTLICEVPAQPS-----PQIHWM 272

Qy   157 LGLLVSHSSSYYFVPEPSDLQSAVSILALTPQSNGTLITCVATWKSLKARKSATVNLTVIRC 216
           ||| :   :: | | | | |: :| : :| | :|:|
Db   273 KD----------GVPLPLPPSPVLILPEIGPQDOGTYSCVATHSSHGPQESRAVSISTIE- 322

Qy   217 PQDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLT------PTCTLTIRCCCC 269
           | |||:||:| | |:||| :| ||| | | || :|:: | | ||:|||
Db   323 PGEEG--------------PTAGSVGGSGLGTLALALGILGGLGTAALLIGVILW 363

Qy   270 RRRCCGCNCCCRCCFCCRRKRGFRIQFQKKSEKEKT--NKETETESGNENSGYNSDEQKT 327
           :|: |: :||:||||
Db   364 QRR-----------------QRRGEERKAPENQEEEERAELN----------- 389

Qy   328 TETASLPPKSCESSSDPEQRNSSCGPP 353
           :| :|| |:||: |||
Db   390 ----------QSEEPEAGESSTGP 404

RESULT 12
AAU77543
ID   AAU77543 standard; Protein; 404 AA.
XX
AC   AAU77543;
XX
DT   05-JUN-2002 (first entry)
XX
DE   Human receptor for advanced glycosylation end product (RAGE).
XX
KW   Receptor for advanced glycosylation end product; RAGE; receptor;
KW   amyloid beta peptide; blood-brain barrier; neurovascular stress;

KW   cerebral vasoconstriction suppressor; cerebral blood flow enhancer;
KW   cerebral amyloid angiopathy; transgenic animal; amyloid beta precursor;
KW   Alzheimer's disease; Down's syndrome; head trauma; stroke; human.
XX
OS   Homo sapiens.
XX
PN   WO200214519-A1.
XX
PD   21-FEB-2002.
XX
PF   14-AUG-2001; 2001WO-US25416.
XX
PR   14-AUG-2000; 2000US-0638648.
XX
PA   (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI   Stern DM, Schmidt AM, Yan SD, Zlokovic B;
XX
PT   Ameliorating neurovascular stress and decreasing cerebral
PT   vasoconstriction in subject suffering from chronic/acute cerebral
PT   amyloid angiopathy, by administering inhibitor of receptor for advanced
PT   glycation endproduct.
XX
PS   Disclosure; Page 16; 68pp; English.
XX
CC   The invention describes a method of ameliorating neurovascular stress,
CC   and decreasing cerebral vasoconstriction in subject suffering from
CC   chronic or acute cerebral amyloid angiopathy, comprising administering
CC   an inhibitor (I) of receptor for advanced glycation end product (RAGE).
CC   (I) inhibits transcytosis of amyloid beta peptides across blood-brain
CC   barrier, thus decreasing cerebral vasoconstriction and increasing
CC   cerebral blood flow. (I) is useful for treating amyloid angiopathy in a
CC   subject, decreasing cerebral vasoconstriction in a transgenic non-human
CC   animal (preferably, transgenic mouse overexpressing mutant human amyloid
CC   beta precursor protein) or a human, suffering from chronic or acute
CC   cerebral amyloid angiopathy, preferably, Alzheimer's disease, and for
CC   ameliorating neurovascular stress comprising cerebral amyloid angiopathy
CC   in a subject, where the neurovascular stress is caused by Alzheimer's
CC   disease, aging, Down's syndrome, head trauma or stroke. This is the
CC   amino acid sequence of human receptor for advanced glycation end
CC   product (RAGE) described in the invention.
XX
SQ   Sequence   404 AA;

Query Match   7.38;   Score 151; DB 23;   Length 404;
Best Local Similarity   23.3%;   Pred. No. 0.00035;
Matches   90; Conservative 38; Mismatches 112; Indels 146; Gaps 19;

Qy   9 PEAVGSGSGNEVIEGPQNATVLKGSQARFNCTVSQG---WKLIMWALSDMVVLSVRPMEP 65
          | || ||||| || |||| | || ||||||||||| || || | || || ||
Db   124 PEIVDSAS--ELTAGVPN-----------------------KVGTCVSEGSYPAGTLSWHLDG-----KP 163

Qy   66 IITNDRFTS------QRYDQGGNFT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGSAY 114
          :|:| | :|:: | |: ||:| | ||| | |:|| |:|| | |
Db   164 LVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRAL 220

Qy   115 LTVQVMGELFIP----SVNLVVAENEP---------------CEVTCLPSHWTRLPDISWE 156
          | | ||||| |||||| ||| ||| ||||||||||||||
Db   221 RTAPIQPRVWEPVPLEEVQLVV----EPEGGAVAPGGTVTLTCEVPAQPS----PQIHWM 272

Qy   157 LGLLVSHSSYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLRARKSATVNLIVIRC 216
          | |||:||:: ||||| :|| | || | ||||||| || |:|| |||||| ||
Db   273 KD--------GVPLPLPSPVLILPEIGPQDQGTYSCVATHSSHGPQESRAVSISIIE- 322

Qy   217 PQDFGGGIINIPGVLSSLPFSLGFSLPTWGKVGLGLAGTMLLT------PTCTLTIRCCCG 269
          | :|||| || ||||::|| | | | | ::| |:|||:| || |:||||
Db   323 PGEEG-----PTAGSVGGSGLGTLIALGILGGIGTRAALLIGVILW 363

Qy   270 RRRCCGCNCCRCCFCCRRKRGFRIQFQKKSEKEKT---NKETETESGNENSGYNSDEQKT 327
          :|:| :|:| |:| |:|
Db   364 QRR--QRRGEERKAPENQEEEEERAELN---------- 389

Qy   328 TETASLPPKSCESSDPEQRNSSCGPP 353
          ||:|| ||:|| ||:|
Db   390 ------------------QSEEPEAGESSTGGP 404

RESULT 13
ID   AAM48745 standard; protein; 404 AA.
XX
AC   AAM48745;
XX
DT   02-APR-2002 (first entry)
XX
DE   Human RAGE protein SEQ ID NO 1.
XX
KW   Human; RAGE; receptor for advanced glycated endproduct; receptor;
KW   antidiabetic; neuroprotective; cytostatic; antiinflammatory; vasotropic;
KW   nephrotropic; dermatological; antiarteriosclerotic; nootropic; diabetes;
KW   Alzheimer's disease; cancer; inflammation; kidney failure;
KW   systemic lupus; nephritis; erectile dysfunction; atherosclerosis.
XX
OS   Homo sapiens.
XX
PN   WO200192892-A2.
XX
PD   06-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US17447.
XX
PR   30-MAY-2000; 2000US-207342P.
PR   05-MAR-2001; 2001US-0799152.
XX
PA   (TRAN-) TRANS TECH PHARMA.
XX
PI   Shahbaz M;
XX
PT   WPI; 2002-114372/15.
XX
DR   Claim 1; Fig 2; 49pp; English.
XX
CC   The invention relates to detecting receptor for advanced glycated
CC   endproducts (RAGE) modulators comprises determining the amount of RAGE
CC   protein or its fragment bound to the pre-adsorbed ligand by measuring the
CC   amount of anti-RAGE antibody bound to the solid surface. The method is
CC   useful for rapid, high-throughput identification of compounds that
CC   modulate RAGE. The compounds are useful for treating symptoms of diabetes
CC   and symptoms of diabetic late complications, amyloidoses, Alzheimer's
CC   disease, cancer, inflammation, kidney failure, systemic lupus nephritis
CC   or inflammatory lupus nephritis, erectile dysfunction and
CC   atherosclerosis.
XX
PT   Detecting a receptor for advanced glycated endproducts (RAGE)
PT   modulators, for treating e.g., cancer, diabetes or inflammation,
PT   comprises measuring the amount of bound anti-RAGE antibody -
XX
PS   ...
XX
SQ   Sequence   404 AA;

Query Match   7.38;   Score 151; DB 23;   Length 404;
Best Local Similarity   23.3%;   Pred. No. 0.00035;
Matches   90; Conservative 38; Mismatches 112; Indels 146; Gaps 19;

Qy   9 PEAVGSGSGNEVIEGPQNATVLKGSQARFNCTVSQG---WKLIMWALSDMVVLSVRPMEP 65
          | || ||||| || |||| | || ||||||||||| || || | || || ||
Db   124 PEIVDSAS--ELTAGVPN-----------------------KVGTCVSEGSYPAGTLSWHLDG-----KP 163

Qy   66 IITNDRFTS------QRYDQGGNFT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGSAY 114
          :|:| | :|:: | |: ||:| | ||| | |:|| |:|| | |
Db   164 LVPNEKGVSVKEQTRRHPETGLFTLQSELM---VTPARGGDPRPTFSCSFSPGLPRHRAL 220

Qy   115 LTVQVMGELFIP----SVNLVVAENEP---------------CEVTCLPSHWTRLPDISWE 156
          | | ||||| |||||| ||| ||| ||||||||||||||
Db   221 ...

RESULT 14

ID   AAW81030 standard; Protein; 1496 AA.
XX
AC   AAW81030;
XX
DT   10-MAY-1999 (first entry)
XX
DE   Melanoma associated antigen MG50.
XX
KW   MG50; melanoma gene-50; melanoma associated antigen; human;
KW   T cell epitope; cancer; lung cancer; rhabdomyosarcoma; diagnosis;
KW   therapy; vaccine.
XX
OS   Homo sapiens.
XX
XX
PN   WO985133-A1.
XX
PD   10-DEC-1998.
XX
PF   04-JUN-1998;   98WO-US11533.
XX
PR   06-JUN-1997;   97US-0870941.
XX
PA   (REGC) UNIV CALIFORNIA.
PA   (UYSC-) UNIV SOUTHERN CALIFORNIA.
XX
PI   Deans RJ, Kan-Mitchell J, Minev BR, Mitchell MS;
XX
DR   WPI; 1999-080820/07.
DR   N-PSDB; AAV99922.
XX
PS   Claim 1; Page 45-49; 79pp; English.
XX
CC   This polypeptide comprises a portion of a new human melanoma
CC   associated antigen, designated MG50. The amino acid sequence was
CC   deduced from a cDNA clone (see AAV99922) isolated from melanoma
CC   cell line MSM M-1 cDNA by subtractive hybridisation. The 5' region
CC   of MG50 cDNA was not obtained. MG50 mRNA has been detected in
CC   melanoma, lung carcinoma and rhabdomyosarcoma cells, foetal brain,
CC   foetal heart and human placenta. The invention also provides T
CC   cell epitopes (see AAW81031-54) from MG50, including cytotoxic and
CC   helper T cell epitopes, antibodies that specifically bind to MG50
CC   or an MG50 T cell epitope, recombinant vectors, and antigen
CC   presenting cells. Methods are provided for identifying an MG50
CC   melanoma associated antigen in an individual and for identifying an
CC   immune response against an MG50 melanoma associated antigen, as
CC   well as methods of stimulating T lymphocytes that are reactive
CC   against cancer cells expressing MG50 and for treating an individual

RESULT 15

ID   AAY70469 standard; Protein; 1496 AA.
XX
AC   AAY70469;
XX
DT   21-JUN-2000 (first entry)
XX
DE   Human p53 target molecule, PRG2 protein.
XX
KW   PRG2; p53 target; human; modulator; cell proliferation; immunomodulatory;
KW   chromosome 2p24.3; cytostatic; gene therapy; tumour cell; inducer;
KW   diagnosis; therapeutic; proliferative disease; cell cycle arrest; cancer;
KW   treatment; apoptosis; knockout animal; cancer susceptibility; dPxn; hPxn;
KW   peroxidasin; redox-regulation; reactive oxygen species; ROS.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   Peptide   1..29
FT   /label= signal_peptide
FT   Protein   30..1496
FT   /label= Mature_human_PRG2_protein
FT   /note= "Homologue of Drosophila peroxidasin, dPxn"
XX
XX
PN   WO200012526-A1.
XX
PD   09-MAR-2000.
XX
PF   27-AUG-1999;   99WO-US19551.
XX
PR   28-AUG-1998;   98US-0098251.
XX
PA   (UYPR-) UNIV PRINCETON.
XX
PI   Horikoshi N, Shenk T;
XX
DR   WPI; 2000-246724/21.
DR   N-PSDB; AAZ51671.
XX
PT   New p53-inducible isolated nucleic acid molecule including open reading

PT   frame encoding human homolog of Drosophila melanogaster peroxidasin,
PT   useful e.g. in detection and treatment of cancer -
XX
PS   Disclosure; Page 71-74; 83pp; English.
XX
CC   The present sequence is the protein encoded by PRG2 gene, that is
CC   upregulated in response to induction of p53 activity in human colon
CC   cancer EB1 cells. This sequence is the human homologue (hPxn) of
CC   Drosophila peroxidasin gene dpxn, that is expressed in heart, placenta,
CC   spleen, ovary and intestines. PRG2 is involved in p53-mediated growth
CC   suppression pathways and plays a role in redox regulation. It is a
CC   haem-peroxidase that increases the intracellular content of reactive
CC   oxygen species (ROS). They are potential targets of p53 regulatory
CC   activity and are useful for modulation of cellular proliferation. PRG2
CC   gene is localised to human chromosome 2p24.3. The PRG target molecules
CC   have cytostatic and immunomodulatory activity. PRG polynucleotides,
CC   proteins and antibodies are useful as diagnostic and therapeutic agents
CC   for detection and treatment of cancer and other proliferative diseases.
CC   The gene/cDNA may be used for gene therapy, to restore a gene function
CC   downstream of p53, that cannot be activated in the p53-deficient tumour
CC   cell. Antibodies can be used as inducers of cell cycle arrest and/or
CC   apoptosis. The DNA sequences can be used to generate 'knockout' animals
CC   as a model of cancer susceptibility.
XX
SQ   Sequence   1496 AA;

Query Match   6.5%;   Score 135;   DB 21;   Length 1496;
Best Local Similarity 25.0%;   Pred. No. 0.041;
Matches   73;   Conservative 36;   Mismatches 127;   Indels   56;   Gaps   15;

Qy   13 GSGSGNEVIEGPQNATVLKGSQARFNCTVS-QGWKLIMWALSDMVVLSVRPMEPIITNDR 71
               ||| ||| | | :   : : | || ||| | |: ||
Db   354 GSPARPTFVIQPQNIEVLVGESVTLECSATGHPPPRISWTRGDRTPLPVDPRVNITPS-- 411

Qy   72 FTSQRVDQGGNFTSEMIIHHVEPSDSGGNIRCSLQNS---RLHGSAYLTVQVMGELFIPSVN 129
               | | ||| | | || : ||||:|:: || :|| : | :|
Db   412 --------GG----LYIQNVVQGDSGEYACSATNNIDSVHATAFIIVQALPQFTVTPQD 458

Qy   130 LVVAENEPCEVTCLPSHWTRLPDISW-ELGLLVSHSSYFVPEPSDLQSAVSILALTPQS 188
                   | | | | |:| || : | | :: | | || || :| | |
Db   459 RVVIEGQTVDFQC-EAKGNPPVIAWTKGGSQLSVDRHLVLSSGTLR--ISGVALHDQ- 514

Qy   189 NGTLTCVATWKSLKARKSATVNLTV------------IRCPQDT---GGGINIPGVLSSLPSL 236
                  | || ||||| |||| || | |||
Db   515 -GQYECQAV--NIIGSQKVVAHLTVQPRVPVFASIPSDTTVEVGANVQLP----CSSQ 566

Qy   237 GFSLP--TWGKVGLGL--AGTMLLTPTCTLTI-----------RCCCCRRCCG 275
                | | || | :: |: ||| ::|||| :| | |
Db   567 GEPEPAITWNKDGVQVTESGKFHISPEGFLIINDVGPADAGRYECVARNTIG 618

XX

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 23:25:02 ; Search time 1905.64 Seconds
                 (without alignments)
                 9985.963 Million cell updates/sec

Title: US-09-729-264-1
Perfect score: 1175
Sequence: 1 ctgtctgccatctgaataa.........gtaatacaactgtcagtatag 1175

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   EST:*
                 1: em_estba:*
                 2: em_esthum:*
                 3: em_estin:*
                 4: em_estmu:*
                 5: em_estcv:*
                 6: em_estpl:*
                 7: em_estro:*
                 8: em_estro:*
                 9: em_htc:*
                10: em_est1:*
                11: gb_est2:*
                12: gb_htc:*
                13: gb_est3:*
                14: gb_est4:*
                15: gb_est5:*
                16: em_estfun:*
                17: em_estom:*
                18: gb_gss:*
                19: em_gss_hum:*
                20: em_gss_inv:*
                21: em_gss_pln:*
                22: em_gss_vrt:*
                23: em_gss_fun:*
                24: em_gss_mam:*
                25: em_gss_mus:*
                26: em_gss_other:*
                27: em_gss_pro:*
                    em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	308.6	26.3	784	12	BG863468	BG863468 60279694T
2	279	23.7	794	12	BF784177	BF784177 60210803 9
3	267.4	22.8	725	12	BG740428	BG740428 60263381 7
4	259.4	22.1	631	10	BB663870	BB663870 BB663870
5	247.8	21.1	527	10	BE032610	BE032610 132035 MA
6	199	16.9	517	10	BE476432	BE476432 159670 BA

RESULT 1
BG863468
LOCUS   BG863468   784 bp   mRNA   linear   EST 29-MAY-2001
DEFINITION 60279694T1F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4917803 5',
            mRNA sequence.
ACCESSION   BG863468
VERSION   BG863468.1 GI:14214006
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 784)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10830 row: b column: 12
            High quality sequence stop: 768.

FEATURES   Location/Qualifiers

ALIGNMENTS

7	191.6	16.3	524	10	AV601192	AV601192 AV601192
8	190.8	16.2	622	17	AG069679	AG069679 Pan trogl
9	190.8	16.2	677	17	AG107877	AG107877 Pan trogl
10	190.8	16.2	471	9	AA265274	AA265274 mx91d09.r
11	186	15.8	611	10	AW990468	AW990468 uf37d05.y
12	186	15.8	931	12	BG173684	BG173684 60233682 3
13	186	15.8	1552	11	AK008060	AK008060 Mus muscu
14	185	15.7	486	10	AI425363	AI425363 mx91d09.y
15	184.4	15.7	663	10	BE376590	BE376590 6C1226419
16	182.8	15.6	650	10	BG206666	BG206666 RST26117
17	168.2	14.3	600	17	AZ379623	AZ379623 uk56d01.y
18	153.2	13.0	216	12	BB625217	BB625217 BB625217
19	153.2	13.0	556	9	AI788300	AI788300 uk56d01.y
20	151.6	12.9	530	9	AI790785	AI790785 uk28a12.y
21	145.8	12.4	423	10	BB846577	BB846577 BB846577
22	143.8	12.2	291	17	AZ411779	AZ411779 M018SB04
23	128.8	11.0	471	12	BF040046	BF040046 BP250004B
24	127.4	10.8	394	10	BB846133	BB846133 BB846133
25	124	10.6	685	17	AG142221	AG142221 Pan trogl
26	110	9.4	785	13	BI452873	BI452873 Pan trogl
27	102.8	8.7	889	13	BI454276	BI454276 60317020
28	100	8.5	185	14	N47851	N47851 yw95h05.r1
29	100	8.5	234	14	N93995	N93995 za66f09.r1
30	98.4	8.4	525	17	AQ403719	AQ403719 HS_5049_A
31	78.6	6.7	360	9	AI647601	AI647601 uk41f12.x
32	78.6	6.7	374	9	AI648720	AI648720 uk28a12.x
33	78.6	6.7	406	9	AI788973	AI788973 uk56d01.x
34	77	6.6	644	10	BB531388	BB531388 BB531388
35	67.8	5.8	397	10	AW112084	AW112084 MC7851 mo
36	63.6	5.4	282	10	BB564363	BB564363 BB564363
37	56	4.8	818	12	BG445021	BG445021 GA_Ea002
38	52.6	4.5	422	9	AI756062	AI756062 EteSTea09
39	51	4.3	457	9	AJ003343	AJ003343 AJ003343
40	50.8	4.3	458	9	AL514929	AL514929 AL514929
41	50	4.3	447	10	AV431233	AV431233 AV431233
42	49.8	4.2	428	10	AA695064	AA695064 GM02059.5
43	49.6	4.2	925	17	CNS038lE	CNS038lE
44	49.6	4.2	942	17	CNS00601	CNS00601 Tetraodon
45	49.4	4.2	481	10	BE017134	BE017134 fk76f06.y

```
source          1..784
                /organism="Mus musculus"
                /strain="NMRI"
                /db_xref="taxon:10090"
                /clone="IMAGE:491803"
                /clone_lib="NCI_CGAP_Mam4"
                /tissue_type="tumor gross tissue"
                /dev_stage="5 months"
                /lab_host="DH10B"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Library constructed by Life Technologies. Investigators
                providing samples: Lothar Hennighausen/Priscilla Furth,
                NIH Reference for transgenic model: Li et al., Cell Growth
                and Differentiation 7, 3-11 (1996)."

BASE COUNT      188 a     202 c     208 g     186 t
ORIGIN

Query Match             26.3%;  Score 308.6;  DB 12;  Length 784;
Best Local Similarity   69.5%;  Pred. No. 2.3e-61;
Matches  447;  Conservative   0;  Mismatches 194;  Indels   2;  Gaps   2;

Qy   24 GAGATGGGGCCTTGTGATTTTCCTCCACGGTTCGTGGGTTCTGGTAATGAAGTCATAGAAGGC 83
Db  123 GTGCTGGTCATCCTGGCTCACGTCACAGCTTCCGGATCCACAGCTTATCAGATCATAGAAGGT 182

Qy   84 CCCCAGAATGCAACAGTGCTGAAGGGCTCCCAGGCTGCCTTCAACTGCACCGTCTCCCAG 143
Db  183 CCTCAGAATGTAACAGTCCTAAAGGACTCAGAGGCTCACTGACGTGACTCAC 242

Qy  144 GGCTGGGAAGCTCATCATGTTGGGCGTCTCCAGTGACATGTGTGTCAGGCCCATG 203
Db  243 GGCTGGGAAGCTTCTCATGTGGACTCTTAACCAAATGGTGGTGCTCACCACCAA 302

Qy  204 GAGCCCATCATCACCAATGACCGCGCTTCACCCTCTCAGAGGTACGACCCAGGGGGGAACTTC 263
Db  303 GGACCCATCATCACCAACAACCGCGCTTCACCTATGCCAGTTACAACAGCACTGACAGCTTC 362

Qy  264 ACCTCGGAGATGAATGATCATCACCAACAATGTGGAGCCCAGTGATTCGGGGAACATCAGATGCCAGC 323
Db  363 ATCTCGGAGTTGAATCATGTCCAGCCCAGTGACTCGGGATCCGTGCAAATCCGTAAGCCAT 422

Qy  324 CTCCAGAACAGAGTCGCCTGCCATGGGTTTGGGATCTGCCAGTTATGGGGAGACTG 383
Db  423 CTGCAGAACAGCCATGGGTTTGGGATCTGCCTTACCGTCCAAGTTATGGGGAGAGCTG 482

Qy  384 TTCATTCCCAGTGTTAATCTTGTAGTCGCCTAGAGAATGAACCTTGTGAAGTTACTTGTCTA 443
Db  483 AACATTCCTAGCAACAACCTTATAGTCACTGAGGGTGAACCCTGTAATGTGACTTGCTAT 542

Qy  444 CCCTCACACTGGACCCGGCTCCCGGAGATAATTCCCGGAGGCTCGGGTCCTCCTGGTTCAGCCAT 503
Db  543 GCCGTGGGTCTGGACCTCCACTCCCGGATAATTTCCTGGGAGGTGAACCCTGAATGTGACCTTACTTGCTAT 602

Qy  504 TCAAGCTATTATTTTGTTCCGGAGCCCAGCGACCTTCAAAGTGCACTGAGCATCCTGGCT 563
Db  603 TCGAGTTACAATTCCTTTCTGGAGCCGGGCCAACTTTATGAGGGTGTCCTGGAC 662

Qy  564 CTGACCCCACACAGAGCAATGGGACTTTGACTTGGCTGGCTACCTGGAAGAGCCTGAAGGCC 623
Db  663 CTCACAACA-TGGGCAACACGGACCTTGACTTGTGTGGCAGAGACTGAAGGAGAGCTG 721

Qy  624 CGCAAGTTCGTCAAACTGTAAAATCTCACTGTGATTCGGTTGTCTCCCC 666
Db  722 AGCAAGTCCTTAACTGTCAA-CTGACTGTCGGTTCAGCTTCCAC 763
```

```
VERSION     BF784177.1  GI:12089213
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 794)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9844 row: n  column: 10
            High quality sequence stop: 694.

FEATURES             Location/Qualifiers
     source          1..794
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4236489"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"

BASE COUNT      186 a     214 c     215 g     179 t
ORIGIN

Query Match             23.7%;  Score 279;  DB 12;  Length 794;
Best Local Similarity   68.3%;  Pred. No. 1.8e-54;
Matches  416;  Conservative   0;  Mismatches 190;  Indels   3;  Gaps   2;

Qy   24 GAGATGGGGCCTTGTGATTTTCCTCCACGGTTCGTGGGTTCTGGTAATGAAGTCATAGAAGGC 83
Db  123 GTGCTGGTCATCCTGGCTCACGTCACAGCTTCCGGATCCACAGCTTATCAGATCATAGAAGGT 182

Qy   84 CCCCAGAATGCAACAGTGCTGAAGGGCTCCCAGGCTGCCTTCAACTGCACCGTCTCCCAG 143
Db  183 CCTCAGAATGTAACAGTCCTAAAGGACTCAGAGGCTCACTGACGTGACTCAC 242

Qy  144 GGCTGGGAAGCTCATCATGTTGGGCGTCTCCAGTGACATGTGTGTCAGGCCCATG 203
Db  243 GGCTGGGAAGCTTCTCATGTGGACTCTTAACCAAATGGTGGTGCTCACCACCAA 302

Qy  204 GAGCCCATCATCACCAATGACCGCGCTTCACCCTCTCACCTCTCAGAGGTGTGGTAGTGTGCTCAGGCCCATG 263
Db  303 GGACCCATCATCACCAACAACCGCGCTTCACCTATGCCAGTTACAACAGCACTGACAGCTTC 362

Qy  264 ACCTCGGAGATGAATGACCGCGCTTCACCTCTCACCAACAATGTGGAGCCCAGTGATTCGGGATCCGTGCAAATCCGTAAGCCAT 323
Db  363 ATCTCGGAGTTGAATCATGTCCAGCCCAGTGACTCGGGATCCGTGCAAATCCGTAAGCCAT 422

Qy  324 CTCCAGAACAGAGTCGCCTGCAATGGGATGCTCGGGCCTGCCTGCAGTTATGGGGAGAGCTG 383
Db  423 CTGCAGAACAGCCATGGGTTTGGGATCTGCCTTACCGTCCAAGTTATGGGGAGAGCTG 482

Qy  384 TTCATTCCCAGTGTTAATCTTGTAGTCGCTGAGAATGAGCCTTGTGAAGTTATTGTGAATCATGTGAACTTGCTAT 443
Db  483 AACATTCCTAGCAACAACCTTATAGTCACTGAGGGTGAACCCTGTAATGTGACTTGCTAT 542

Qy  444 CCCTCACACTGGACCCGGCTCCCGGAGATAATTTCCTGGGAGGTGAACCCTGTAATGTGACTTACTTGCTAT 503
Db  543 GCCGTGGGTCTGGACCTCCACTCCCGGATAATTTCCTGGGAGGTGAACCCTGAATGTGACCTTACTTGCTAT 602

Qy  504 TCAAGCTATTATTTTGTTCCGGAGCCCAGCGACCTTCAAAGTGCACTGAGCATCCTGGCT 563
```

```
Db   603  TCGAGTTACAATTCCTTTCTGGAGCC--GGGCACTTTATGAGGGGTGGAGCC
          ||||| |||| || |  ||||||||||    ||||| ||| ||  | |||||||
Qy   564  CTGACCCACAGAGACCAATGGGACTTTGACTTGCGTGCGTGCCTACCTGGAAGACAGAGCCTTGAAGGCC 623
          |||   |||| ||  ||||| ||||| || |||| |||  |   |||| ||| |||| ||| || ||| |
Db   661  CTCA-CACACTGGGCAACGGGACCTTGACTTGTGTGCCAAGAGGCTGAGGCACTTGCAGGC 660
Qy   624  CGGCAAGTCT 719
          |  |||||||
Db   720  CAGCAGTCT 728
```

RESULT 3
BG740428

```
LOCUS       BG740428                725 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602633817F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4778789 5',
            mRNA sequence.
ACCESSION   BG740428
VERSION     BG740428.1  GI:14051081
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 725)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov.
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov/.
            Plate: LLAM1635  row: b  column: 06
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1..725
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4778789"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site.1: NotI;
                     Site.2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP library."
BASE COUNT      227 a    163 c    159 g    176 t
ORIGIN
```

```
Query Match     22.8%;  Score 267.4;  DB 12;  Length 725;
Best Local Similarity   99.6%;
Matches  268;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   906  AAGAAAATTCTGAAAAAGAGAACAAAACAAAACAAACAACTGAGACAGAAGTGGAAATGAAAAC 965
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   377  AGGAAAATCTGAAAAAGAGAACAAAACAAAACAAACAACTGAGACAGAAGTGGAAATGAAAAC 436

Qy   966  TCCGGCTACAATTCAGATGAACAAAAAGACCACAGACACCGGTTCTCTCCCCCCAAATCC 1025
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   437  TCCGGCTACAATTCAGATGAACAAAAAGACCACAGACACCGGTTCTCTCCCCCCAAATCC 496

Qy  1026  TGTGAATCCAGTGATCCTGAACAAACAAGAAAAACACAGACACCGGTTCTCTCCCCCTCACCACCACGCGGGCT 1085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   497  TGTGAATCCAGTGATCCTGAACAAACAAGAAAAACAGTAGCTGTGGCCCTCCTCACCACCAGTCCCC 556

Qy  1086  GATCAACGTCCACCCACAGGCCAGCAAGTCATCCACCAGGGCTTCTTTAATCTGGCCCAGTCCT 1145
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   557  GATCAACGTCCACCCACAGGCCAGCAAGTCATCCACCAGGGCTTCTTTAATCTGGCCCAGTCCT 616
```

```
Qy  1146  GAGGAAGGTCAGTAATACAACTGTAGTATA 1174
          ||||||||||||||||||||||||||||||
Db   617  GAGGAAGGTCAGTAATACAACTGTAGTATA 645
```

RESULT 4
BB663870

```
LOCUS       BB663870                631 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB663870 RIKEN full-length enriched, 0 day neonate lung Mus
            musculus cDNA clone E030016M12 5', mRNA sequence.
ACCESSION   BB663870
VERSION     BB663870.1  GI:16497624
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 631)
AUTHORS     Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            ,M., Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL     Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res.  10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohata,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S. Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res.  11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
            e mouse tissues.
FEATURES             Location/Qualifiers
     source          1..631
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="E030016M12"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     lung"
                     /tissue_type="lung"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
```

COMMENT Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAAGCGGCCGCAACTCCGAGTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAAGGTTCGAGTTAAATTAAATTAAATCCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

```
BASE COUNT      154 a     167 c     161 g     149 t
ORIGIN
```

RESULT 5

BE032610

```
Query Match      22.18;   Score 259.4;  DB 10;   Length 631;
Best Local Similarity   70.4%;           Pred. No. 6.4e-50;
Matches   347; Conservative   0; Mismatches 146; Indels   0;  Gaps   0;

Qy   24 GAGATGGGGCGTTGTGATTTTCCTCCACGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGC   83
         ||||||| |||||||||||||||| ||||||||||||||||||| |||||||||||||||
Db  136 GTGCTGGTCATCCTGGCCTCAGCTGACAGCTTCGGGATCCACCGTTCTCCCAG 195

Qy   84 CCCCAGAAATGCAAACAGTCGCTTGGAACGTCGCGTTCACCGTCCACCGTCTCCCAG  143
         |||| || |||| ||||||| || |||||||||||||||||||||| | |||||
Db  196 CCTCAGAATGTAACAGTCCTAAAGGACTCAGAGGCTCACTCAACTGCACCGTGACTCAC  255

Qy  144 GGCTGGAAGCTCATCAATTGTGGGCTCTCCAGTCACATGGTTGGTTGCTAAGCGTCAGGCCCATG  203
         ||||||||||| |||| ||||| |||| |  |    |||||  ||||||| |  |||||  || ||
Db  256 GGCTTGGAAGCTTCATTGTGGACTCTTTTAACCCAAAATGGGTGGTTGCTCGAGTGTTCTCACACCAA  315

Qy  204 GAGCCCATCACACAATGACCGGCGTTCACCTTCTTCAGAGGGTACGACCCAGGGGGGGGAACTTC  263
         |||||||||| || |||| |  ||||||| | ||||    | ||||||| || |||| |||||||
Db  316 GGACCCATCATCACCAACAACACCGCTTCACCTATGCCAGTTACACAACAGCACTGACAGCTTC  375

Qy  264 ACCTTCGGAGATGATCATCCACACAAATGTGGGAGCCCAGCCAGACACATCAGATGCAGC  323
         |  | ||||| || |||||||| ||||| |||||| |||  ||| |||| ||||||||||
Db  376 ATCTCGGGAGTTGATCATCCATCAAGATGATGTGCAGCGCCCAGGGATCCGTGCAATGCAGC  435

Qy  324 CTCCAGAACAGTCGCCTGCATGGATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTG  383
         || ||||||| | | ||||| ||| |||  ||||| |||  |||| |||||| ||| |||
Db  436 CTGCAGAACAGCCATGGGTGTTGGCATCTGCCTTCTCTCCAGTCAGTCATAGGGGACCCTG  495

Qy  384 TTCATTCCCAGTGTTAAATCTTGTTGTCGGCTGAGAAGTGAAGTTGTTGTGCTA  443
         |||||||| || |||||  | |  |  |   |||  |||||||| | ||  ||||
Db  496 AACATTCCTAGCAACAACCTTATAGTCACTGAGGGTGAACCCTGTAATGTGACTTGCTAT  555

Qy  444 CCCTCACACTGGACCCGGCCCCGGGATTATTCCTGGGAGCGTCGGGTCCGGTCAGCCAT  503
         | | |||| |||||| | |   |||| | |||||||||||   | ||| ||| ||||||
Db  556 GCCGGTGGGCTGGACCTCACTCCCGGAТATTTCCTGGGGAGCTTGAGGTTCCGGTAAGCCAT  615

Qy  504 TCAAGCTATTATT  516
         || || | | |||
Db  616 TCGAGTTTACAATT  628
```

```
LOCUS       BE032610     527 bp   mRNA   linear   EST 09-JUL-2000
DEFINITION  132035 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BE032610
VERSION     BE032610.1   GI:8327619
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 527)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
```

COMMENT EST discovery in swine

JOURNAL Unpublished (2000)
```
            Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAAACAAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 63   Row: E  Column: 13
            Seq primer: ATTTAGGTGACACTATAG.
```

FEATURES   Location/Qualifiers
```
     source          1..527
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."

BASE COUNT      105 a     147 c     176 g      99 t
ORIGIN
```

RESULT 6

BE476432

```
Query Match      21.18;   Score 247.8;  DB 10;   Length 527;
Best Local Similarity   76.7%;           Pred. No. 3.1e-47;
Matches   303; Conservative   0; Mismatches  92; Indels   0;  Gaps   0;

Qy   24 GAGATGGGGCGTTGTGATTTTCCTCCACGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGC   83
         |||||||||||||||||||| ||||||||||||||||||||| ||||||||| |||||||
Db  133 GCGCTGGTCGTCGTTCTGGCCTGGCAGCTTGTGTGGATCCGGCAGTTGTGGAAAACTTAGAGGGT  192

Qy   84 CCCCAGAAATGCAAACAGTCCTGAAAGGGCTCCCAGCCAGGCTTGTGTGGAACGTCGCGTTCACCGTCCACCGTCTCCCAG  143
         || | ||||| |||| |||||  || ||| || |  | |   | | |||||||| || |||||
Db  193 CCAGGAAACGTCACGGCCGGTGAAGGGCTCGGAGGCTCGCTTCAACTGCACCATCTCCCAG  252

Qy  144 GGCTGGAAGCTCATCAATGTGGGCTCTCAGTGACAATGGTGGTTGCTAAGCGTCAGGCCCATG  203
         |||| |||||||||||||||||||||| | |  ||| ||| ||||||| | |||||||| ||
Db  253 GGCTTGGAAGCTCATCAATGTGGGCCTCTGAAAAGGCACGGTGGTTCTGAGCGTCAGGCCCACC  312

Qy  204 GAGCCCATCACACAATGACCGGCGTTCACCTCTCAGAGGGTACGACCCAGGGCGCGGAAACTTC  263
         ||||||||| || |||| |  | |||||| | ||| |  ||  |||  |||||| | ||||||
Db  313 GAGCCCATCATCACCAACAACACCGAATTCACGTCTGCAAGTTATGAAGAGAGGCGGGTTAAACTTC  372

Qy  264 ACCTTCGGAGATGATCATCCACACAAATGTGGGAGCCCAGGAACAACATCAGATGCAGC  323
         | || ||||| || |||||||| ||||| ||| || ||||  | || ||||||| |||
Db  373 ATCTCTGAGATGATGATCATCCATCAATGACGTGCAGCGCCCAGGGCTCAAATGCAGC  432

Qy  324 CTCCAGAACAGTCGCCTGCATGGAATCGTGCTTACCGTCCAAGTTATGGGAGAGCTG  383
         |||||||||||| | | |||||||| || || |||||||||||||||| |||| |||
Db  433 CTCCAGAACAGTGACCGGGATGGATCCGCGTTTCCTTTCCTTTCCTTCATGGGGGTGTGGGGGATGCCTG  492

Qy  384 TTCATTCCCAGTGTTAAATCTTGTTGTCGGCTGAGAA  418
         |||||||| ||||| ||| |||  | | |   ||| ||
Db  493 CTCATTCCCGGAGGCAGCCTCGTAGTCCTTCGGGA  527
```

```
LOCUS       BE476432     517 bp   mRNA   linear   EST 28-AUG-2000
DEFINITION  159670 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE476432
VERSION     BE476432.1   GI:9595965
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
```

REFERENCE　1 (bases 1 to 517)
AUTHORS　Sonstegard,T.S., Capuco,A.V., Van Tassell,C.P., Ashwell,M.S. and Wells,K.D.
TITLE　Mapping of Expressed Sequence Tags from a normalized bovine mammary gland cDNA library
JOURNAL　Unpublished (2000)
COMMENT　Contact: Sonstegard TS
USDA, ARS, Beltsville Agricultural Research Center
Bdlg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@anri.barc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 27　row: H　column: 19
Seq primer: ATTTAGGTGACACTATAG.

FEATURES　Location/Qualifiers
source　1..517
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="BARC 5BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled mRNA isolated from mammary
tissues at eight physiological, developmental, and disease
states."

BASE COUNT　117 a　133 c　165 g　102 t
ORIGIN

RESULT 7

LOCUS　AV601192　524 bp　mRNA　linear　EST 27-NOV-2001
DEFINITION　AV601192 Bos taurus kidney fetus Bos taurus cDNA clone EIKI002H04 5', mRNA sequence.
ACCESSION　AV601192
VERSION　AV601192.1　GI:9723510
KEYWORDS　EST.
SOURCE　cow.

Query Match　16.9%;　Score 199;　DB 10;　Length 517;
Best Local Similarity　73.8%;　Pred. No. 7.1e-36;
Matches 253;　Conservative 0;　Mismatches 90;　Indels 0;　Gaps 0;

```
Qy   24 GAGATGGGGCGTTGTGAPTTTTCCTCACACGGTTCTGGGGTTCTGGTAATGAAGTCATAGAAGGC   83
Db  175 GTGCTCGGTGGGTCCTGGGCCGGTGGCAGCCIGTGGATCCAGCAGTGAAAATCATAGAGGGT    234

Qy   84 CCCCAGAATGCAACAGTGTTCCTGAAAGGGGCTCCCAGGCTCGACCGTTCTCTCCCAG        143
Db  235 CCCAAGAANTGTCACAGCCCTGAAAGGGCTCGGAGGCTCGCTTCAACTGCACCCATCTCCCAG   294

Qy  144 GGCTGGAAGCTCATCANTGTGGGCGCTCTCAGTGACACATGGTGGTGCTAAGCGTCAGGGCCATG  203
Db  295 GGCTGGGAAGCTCGTCAATGTGGGCGGCTCTGAGAGGCCACAGTGGTGCTGAGCATGACAACTTAAT 354

Qy  204 GAGCCCATCACICACCAAATGACGGCTTCACCTCTCAGAGGTAGCACCAGGCGGGGAACTTC     263
Db  355 GAGACCCATCACICACCAGTGACGCCTTCACTTCGGCAAGCTACCAAGAGAGGGCGGGAAACTTC  414

Qy  264 ACCTCGGAGATGATGATCATCCACAAATGTGGAGGCCCCAGTGATTCGGGACAACATCAGATGCAGC 323
Db  415 ATCTCTCGAGATGATAATTCATGACGTGACGTGCAACTGCACGATGCCGGGCAAGTCAAATGCAGC  474

Qy  324 CTCCAGAAACAGTCGCCTGCATTGCTTACCTTACCGTCC                           366
Db  475 CTTCAGAACACGCAAATCGGGATGGAGCGCGCGCCTTCCTTTTCTGTTC                 517
```

ORGANISM　Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bovini; Bos.
REFERENCE　1 (bases 1 to 524)
AUTHORS　Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
TITLE　Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
JOURNAL　Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE　21570554
COMMENT　Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocoa.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.

FEATURES　Location/Qualifiers
source　1..524
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bos taurus kidney fetus"
/tissue_type="kidney"
/dev_stage="fetus"
/lab_host="DH10B"
/note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
was deleted from a NotI site"

BASE COUNT　120 a　135 c　168 g　96 t　5 others
ORIGIN

Query Match　16.3%;　Score 191.6;　DB 10;　Length 524;
Best Local Similarity　73.6%;　Pred. No. 3.7e-34;
Matches 242;　Conservative 0;　Mismatches 87;　Indels 0;　Gaps 0;

```
Qy   24 GAGATGGGGCGTTGTGAPTTTTCCTCACACGGTTCTGGGGTTCTGGTAATGAAGTCATAGAAGGC   83
Db  196 GTGCTCGGTGGGTCCTGGGCCGGTGGCAGCCIGTGGATCCAGCAGTGAAAATCATAGAGGGT    255

Qy   84 CCCCAGAATGCAACAGTCCTGAAAGGGGCTCCCAGGCTCGACCGTTCAACTGCACCGTCTCCCAG  143
Db  256 CCCAAGAANTGTCACAGCCCTGAAAGGGCTCGGAGGCTCGCTTCAACTGCACCATCTCCCAG    315

Qy  144 GGCTGGAAGCTCATCANTGTGGGCGCTCTCAGTGACACATGGTGGTGCTAAGCGTCAGGGCCATG  203
Db  316 GGCTGGGAAGCTCGTCAATGTGGGCGGCTCTGAGAGGCCACAGTGGTGCTGAGCATGACAACTTAAT 375

Qy  204 GAGCCCATCACICACCAAATGACGGCTTCACCTCTCAGAGGTACGACCAGGCGGGGAACTTC     263
Db  376 GAGACCCATCACICACCAGTGACGCCTTCACTTCGGCAAGCTACCAAGAGAGGGCGGGAAACTTC  435

Qy  264 ACCTCGGAGATGATGATCATCCACAAATGTGGAGCCCCAGTGATTCGGGAACATCAGATGCAGC   323
Db  436 ATCTNTGAGATGATAATTCATGACGTGACGTGCAACTCACGATGCCGGGCAGTNAAATGCAGC    495

Qy  324 CTCCAGAACAGTCGCCTTGCATTGGGGATCTGC                                 352
Db  496 CTTCANAACAGCAATCGGGATGGAGCCGC                                     524
```

RESULT 8

LOCUS　AG069679　622 bp　DNA　linear　GSS 03-NOV-2001
DEFINITION　Pan troglodytes DNA, clone: PTB-060F12.F, genomic survey sequence.
ACCESSION　AG069679
VERSION　AG069679.1　GI:16621481
KEYWORDS　GSS.
SOURCE　Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC library clone:PTB-060F12.F.

Left column:

```
ORGANISM   Pan troglodytes
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE  1
  AUTHORS  Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
           Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE    BAC end sequences of Library PTB
  JOURNAL  Unpublished
REFERENCE  2 (bases 1 to 622)
  AUTHORS  Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
           Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE    Direct Submission
  JOURNAL  Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
           and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
           1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           (E-mail:chimpbes@gsc.riken.go.jp; URL:http://hgp.gsc.riken.go.jp/,
           Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT    Clones are derived from the chimpanzee BAC library PTB This BAC end
           was generated during the R&D process and may have higher chance of
           clone tracking errors.
FEATURES            Location/Qualifiers
     source         1..622
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="PTB-060F12.F"
                    /sex="male"
                    /cell_type="lymphoblast"
                    /clone_lib="PTB Chimpanzee Male BAC Library"
     PRIMERS
                    Sequencing: -21M13
     LIBRARY
                    Vector    : pKS145
                    R.Site 1  : SacI
                    R.Site 2  : SacI
BASE COUNT      165 a      153 c      128 g      176 t
ORIGIN

Query Match       16.28%;  Score 190.8;  DB 17;  Length 622;
Best Local Similarity  94.3%;  Pred. No. 5.8e-34;
Matches  198;  Conservative 0;  Mismatches 12;  Indels 0;  Gaps 0;

Qy   474 TCCTGGGAGCTCGGTCTCTCCTGGTCAGCCATTCAAGCTATATATTTGTTCCGGAGCCCAGC 533
         ||||| |||||||||||||||||||||||||||||||||| ||| |||||||||||||||||
Db    72 TCCGCGGAGCTCGGTCTCTCCTGGTCAGCCATTCAAGCTATTATTTTGTTCCGGAGCCCAGC 131

Qy   534 GACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGACCCCACAGAGCAATGGGAGACTTTGACT 593
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   132 GACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGACCCCACAGAGCAATGGGAGACTTTGACT 191

Qy   594 TGCGTGGCTACCTGGAAGAGCCTGAAGAGCCCCAAGGTCTGCAACTGTAAATCTCACTGTG 653
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   192 TGCGTGGCTACCTGGAAGAGCCTGAAGAGCCCCAAGGTCTGCAACTGTAAATCTCACTGTG 251

Qy   654 ATTCGGGTGTCCCCAAGACACTGGAGGTGGT 683
         |||||||||||||||||||||       ||
Db   252 ATTCGGGTGTCCCCAAGGTAAGTGAAGATGT 281

RESULT 9
LOCUS       AG107877                677 bp    DNA     linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-112N07.F, genomic survey sequence.
ACCESSION   AG107877
VERSION     AG107877.1  GI:16728395
KEYWORDS    GSS.
SOURCE      Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-112N07.F.
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
```

Right column:

```
  TITLE    BAC end sequences of Library PTB
  JOURNAL  Unpublished
REFERENCE  2 (bases 1 to 677)
  AUTHORS  Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
           Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE    Direct Submission
  JOURNAL  Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
           and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
           1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           (E-mail:chimpbes@gsc.riken.go.jp; URL:http://hgp.gsc.riken.go.jp/,
           Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT    Clones are derived from the chimpanzee BAC library PTB This BAC end
           was generated during the R&D process and may have higher chance of
           clone tracking errors.
FEATURES            Location/Qualifiers
     source         1..677
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"
                    /clone="PTB-112N07.F"
                    /sex="male"
                    /cell_type="lymphoblast"
                    /clone_lib="PTB Chimpanzee Male BAC Library"
     PRIMERS
                    Sequencing: -21M13
     LIBRARY
                    Vector    : pKS145
                    R.Site 1  : SacI
                    R.Site 2  : SacI
BASE COUNT      168 a      159 c      143 g      207 t
ORIGIN

Query Match       16.28%;  Score 190.8;  DB 17;  Length 677;
Best Local Similarity  94.3%;  Pred. No. 5.9e-34;
Matches  198;  Conservative 0;  Mismatches 12;  Indels 0;  Gaps 0;

Qy   474 TCCTGGGAGCTCGGTCTCTCCTGGTCAGCCATTCAAGCTATATATTTGTTCCGGAGCCCAGC 533
         ||||| |||||||||||||||||||||||||||||||||| ||| |||||||||||||||||
Db    72 TCCGCGGAGCTCGGTCTCTCCTGGTCAGCCATTCAAGCTATTATTTTGTTCCGGAGCCCAGC 131

Qy   534 GACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGACCCCACAGAGCAATGGGAGACTTTGACT 593
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   132 GACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGACCCCACAGAGCAATGGGAGACTTTGACT 191

Qy   594 TGCGTGGCTACCTGGAAGAGCCTGAAGAGCCCCAAGGTCTGCAACTGTAAATCTCACTGTG 653
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   192 TGCGTGGCTACCTGGAAGAGCCTGAAGAGCCCCAAGGTCTGCAACTGTAAATCTCACTGTG 251

Qy   654 ATTCGGGTGTCCCCAAGACACTGGAGGTGGT 683
         |||||||||||||||||||||       ||
Db   252 ATTCGGGTGTCCCCAAGGTAAGTGAAGATGT 281

RESULT 10
LOCUS       AA265274                471 bp    mRNA    linear   EST 20-MAR-1997
DEFINITION  mx91d09.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:693713
            5', mRNA sequence.
ACCESSION   AA265274
VERSION     AA265274.1  GI:1901370
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 471)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
```

RESULT 11

```
LOCUS       AW990468
DEFINITION  uf37d05.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
            IMAGE:1513545 5' similar to TR:060962 O60962 L549.1.;, mRNA
            sequence.
ACCESSION   AW990468
VERSION     AW990468.1  GI:8186027
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 611)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:940397
            Seq primer: -40RP from Gibco
            High quality sequence stop: 474.
FEATURES             Location/Qualifiers
     source          1..611
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1513545"
                     /clone_lib="Soares_mammary_gland_NMLMG"
                     /sex="female (lactating)"
                     /tissue_type="mammary_gland_NMLMG"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; 1st strand cDNA was prepared from mammary
                     gland tissue from a lactating female, and was then primed
                     with a Not I - oligo(dT) primer. Double-stranded cDNA was
                     ligated to Eco RI adaptors (Pharmacia), digested with Not
                     I and cloned into the Not I and Eco RI sites of the
                     modified pT7T3 vector. Library is normalized. Library
                     was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      152 a    155 c    156 g    148 t
ORIGIN
```

```
Query Match         15.8%;  Score 186; DB 10;  Length 611;
Best Local Similarity  71.1%;  Pred. No. 7.6e-33;
Matches 246;  Conservative   0;  Mismatches 100;  Indels   0;  Gaps   0;

Qy   24 GAGATGGGGCTTGTGAGTTTCCTCCACGGTTCTGGGTCTTGGGTAATGAAGTCATAGAAGGC 83
        |||||||||||||| ||||  ||  |||| ||| |||||| |||||||||||| |||||||||
Db   24 GAGATGGGGCTTGTGAGTTTCCTCCACGGTTCTGGGTCTTGGGTAATGAAGTCATAGAAGGC 83

Qy   84 CCCAGAATGCAACAGTCCCAGGGCTCCCAGGGTCGCTTCAACGTCACCGTCTCCCAG 143
        |||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   99 GTGCTGGTCATCCTGGCTCGGCCTCCCAGGGTCCCAGGTCCGGATCCAGAACTTCTTGGCTAAATGAAGGT 158
```

```
RESULT 12
LOCUS       BG173684
DEFINITION  602336823F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4459973 5',
            mRNA sequence.
ACCESSION   BG173684
VERSION     BG173684.1  GI:12680387
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1 (bases 1 to 931)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10260 row: n column: 06
            High quality sequence stop: 628.
FEATURES             Location/Qualifiers
     source          1..931
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4459973"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      233 a    277 c    242 g    179 t
ORIGIN

Query Match             15.8%;  Score 186;  DB 12;  Length 931;
Best Local Similarity   71.1%;  Pred. No. 7.9e-33;
Matches   246;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy       24 GAGATGGGGCGTTGATTTTCCTCCACGGTTCTGGGTCTGGTAAGTCATAGAAGGC 83
            || |||||   |||||||||  |||||||| |||  |||| |||| |||||||||
Db      131 GTGCGTGGTCATCCTGGCTCACCTGACAGCTTCCGGATCCAGTTATCAGATCATAGAAGGT 190

Qy       84 CCCCAGAATGCAACAGTCCTGAAAGGCGCTCCCCAGGCTGCTTCACGTCACGTCTCCCAG 143
            |||||||||||||| |||||||| ||||| | |||| |||| || |||| |||||||||
Db      191 CCTCAGAAATGTAACAGTCTTAAAGGACTCAGAGGGCTCACTTCAACTGCACCGTGACTCAC 250

Qy      144 GGCTGGGAAGCTCATCATGTGGGCTCTCAGTGACATGGTGGTGTCTAAGCGTCAGGCCCATG 203
            |||||||| |||||||| || || ||  ||||   ||| |   || ||||| |||  ||
Db      251 GGCTGGGAAGCTTCTCATGTGGACTCTTAACCAAATGGTTGGTGGCTGTCTCACCACCAA 310

Qy      204 GAGCCCATCATCACCAAATGACCGCGTTCACCCTCTCAGAGGTACGACCAGGGGGGGAACTTC 263
            |||||||| |||||||||   ||||| ||||||  |||||||| || ||||||||||  ||
Db      311 GGACCCATCATCACCAACAACCGCTTCACCTTATGCCAGTTACACAACGACCACTGACAGCTTC 370

Qy      264 ACCTCGCGAGATGATCANCATCCACAAATGTGGAGCCCAGTGGAACATCAGATGCCAGC 323
            ||||| ||||| || ||    ||| || |||||||  ||||||| ||||  ||||||||
Db      371 ACTCGGAGTTGATCANCCATCAAGATGTCGCAGCCACCAGTGACTCGGGATCCGTCCAAG 430

Qy      324 CTCCAGAACAGTCGCCTGCACATGGGATCTGCTTACCTTACCGTCCAAG 369
            || |||||||| ||||  ||||||| ||  || ||   ||||||| |
Db      431 CTGCAGAACAGCCATGGGTTTGGATCTGCCTTCCTCTCTCAGTGCCAAG 476

RESULT 13
LOCUS       AK008060                1552 bp    mRNA    linear    HTC 19-JAN-2002
DEFINITION  Mus musculus adult male small intestine cDNA, RIKEN full-length
            enriched library, clone:2010003D20:homolog to IGF5 PROTEIN
            (FRAGMENT), full insert sequence.
ACCESSION   AK008060
VERSION     AK008060.1  GI:12842009
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (strain:C57BL/6J) adult male small intestine cDNA to
            mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library
            clone:2010003D20.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
   PUBMED   10349636
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system-384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojoborl,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5 (bases 1 to 1552)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,H., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamanura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN) Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
```

```
COMMENT     RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues. First strand cDNA was primed with a primer
            [5'-GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTVN 3'], cDNA was
            prepared by using trehalose thermo-activated reverse transcriptase
            and subsequently enriched for full-length by cap-trapper. cDNA went
            through one round of normalization to Rot = 5.0 and subtraction to
            Rot = 20.0. Second strand cDNA was prepared with the primer adapter
            of sequence[5'-
            GAGAGAGAGATTCTCAGTTAATTAAATTAAATCCCCCCCCCCCC 3'], cDNA was cleaved
            with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
            Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..1552
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2010003D20"
                     /db_xref="MGD:MGI:1904786"
                     /db_xref="GI:12842010"
                     /db_xref="taxon:10090"
                     /clone="2010003D20"
                     /sex="male"
                     /tissue_type="small intestine"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             73..885
                     /note="data source:SPTR, source key:Q9NSI5, evidence:ISS
                     homolog to IGSF5 PROTEIN (FRAGMENT)
                     putative"
                     /codon_start=1
                     /protein_id="BAB25435.1"
                     /db_xref="MGI:1919308"
                     /translation="MEGSWRDVLAVLGVILAQLTASGSSYQIEGPQNVTVLKIDSEAHF
                     NCTVTHIGWKLLMWTLNQMVVLSLTTQGPIITNNRFTYASYNSYDSFISELIIHDVQPS
                     DSGSVQCSLQNSHGFGSAFLSVQDSIGEEGPALPTWAIILLAVAFSLLILIIVLLII
                     LPPKSAEVSLPEKRSSSLPVQELNKHQPGPATHPRVSFDIASPQKVRNVTLV"
     polyA_signal    1534..1539
                     /note="putative"
     polyA_site      1552
                     /note="putative"
BASE COUNT      430 a    369 c    374 g    379 t
ORIGIN
Query Match             15.8%;  Score 186;  DB 11;  Length 1552;
Best Local Similarity   71.1%;
Matches  246;  Conservative    0;  Mismatches 100;  Indels    0;  Gaps    0;

Qy         24 GAGATGGGGCGTTGTGATTTTCCTCCACGGTTCTGGGTCGGTAATGAAGTCAATAGAAGGC 83
              |||||||||| || || ||||| | |||||| || |||||||  | |||||| |||||||
Db        103 GTGCTGGTCATCCTGGCTCAGCTGACAGCTTCCGGATCCAGTTATCAGATCAATAGAGGGT 162

Qy         84 CCCCAGAATGCAACAGTCCTGAAGGGCTCCCAGGCGCTCGCTTCAACTGCACCGTCCCCAG 143
              || |||||| || |  ||  | ||||||| | ||||||| |  ||||||||| |||||||
Db        163 CCTCAGAATGTAACAGTCCTAAAGGGACTCACTTCAACTGCACCGTCGACTCAC 222

Qy        144 GGCTGGAAGCTCATCATGTGGGGCGCTCAGTGACATGGTGTCGCTAAGGCTCAGGCCCATG 203
              |||||| || | || |  ||||||| | |||| | || |   || || ||||| |||||
Db        223 GGCTGGAAGCTTCTCATGTGGACTCTTACCAAATGGTGTGGTCTTCACCACCA 282

Qy        204 GAGCCCATCATCACCAATGACCGCGTTCACCTCTCAGAGGTACGACGGGGGGAACTTC 263
              ||||| | | |  ||||  |||||||   | || ||||| ||| ||| |  |  |  |
Db        283 GGACCCATCATCACCAACAACCGCGTTCACCTATGCCCAGCTCAGCAGTACCGACAGCTTC 342

Qy        264 ACCTCGGAGATGATCATCCACAAATGTGGAGCCCAGTGATTCGGGGACATCAGATGCAGC 323
              |  |   | || |  |   |  | ||||||| |  | || | || ||| ||| ||  |||
Db                                                                     
```

```
RESULT 15
LOCUS       BE376590                 663 bp    mRNA      linear   EST 21-JUL-2000
DEFINITION  601226419F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3584645 5',
            mRNA sequence.
ACCESSION   BE376590
VERSION     BE376590.1  GI:9321955
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 663)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8743 row: f  column: 06
            High quality sequence stop: 498.
FEATURES             Location/Qualifiers
     source          1..663
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3584645"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally; Primer: Oligo dT.
                     library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      143 a    193 c    170 g    157 t
ORIGIN
```

Query Match   15.7%;   Score 184.4;   DB 10;   Length 663;
Best Local Similarity   70.8%;   Pred. No. 1.8e-32;
Matches 245;   Conservative   0;   Mismatches 101;   Indels   0;   Gaps   0;

```
Qy     24 GAGAATGGGGGCTTTGTGATTTTCCTCCACGGTTCTGGGTTCTGGTAATGAAGTCATAGAAGCC   83
          | | || || ||| ||||| || || |||  || |||  ||  || ||  ||  | ||||
Db     65 GTGCTGGTCATCCTGGCTCAGCTGAACAGCTTCCGGATCCAGTTATCAGATCAATAGAAGGT  124

Qy     84 CCCGAGAATGCAACAGTCTCTGAAGGGCGTCCCAGGCTCGCTTCAAACTGCACCGTCTCCAG  143
          || |||||| |||  |||||  ||||||| |||| |||  ||||||||||||  || || ||
Db     84 CCCAGAAAATGCAACAGTCTCCGAAGGGCTCCCAGGCTCGCTTCAAACTGCACCGTCTCCAG  143
```

```
Qy    144 GGCTGGAAGCTCATCATGTGGGGCTTCCAGTGACATGGGTGGTGGTGCTTAAGCGTCAGGCCCATG  203
          ||||||||| |||||||||||||| ||||||| || ||||| ||||| || |||| || |||
Db    144 GGCTGGAAGCTTCTCATGTGGGGGCTTCAGTGACATGGGTGGTGGTGCTTAAGCGGTCAGGCCCATG  203

Qy    204 GAGCCCATCACCAATGACCGCTTCACCTCTCAGAGGTACGACCAGGGCGGGGAACTTC  263
          |||||||||||||||| ||||||||||| ||||||| | |||||| ||||||||||
Db    185 GAGCCCATCACCAATGACCGCTTCACCTCTCAGAGGTACGACCAGGGCGGGGAACTTC  244

Qy    264 ACCTCGGAGATGATCATCCACACAAATGTGGAGCCCAGTGATTCGGGGACAATCAGATGCAGC  323
          |||| | |||||| |||| ||  |  || |||||||||||| ||||||||| |||| ||||||
Db    245 GGACCCATCATCACCAACAACCGCTTCACCTATGCCAGTTACAACAGCACTGACAGCTTC  304

Qy    270 GGACCCATCATCACCAACAACCGCTTCACCTATCGCCAGTTACAACAGCACTGACAGCTTC  329
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    305 ATCTCGGAGTTTGATCATCCATGATGTGCAGCCCAGCCCAGTCGGGGATCCGTGCAAATGCAGC  364

Qy    330 ATCTCGGAGTTTGATCATGATGTGCAGCCCAGTGACTCGGGGATCCGTGCAAATGCAGC  389
          ||||||||||||||||| ||||||||||||||| |||||||||||||||||||||||||

Qy    324 CTCCAGAACAGTCGCCTGCATGGATCTGCTTACCTTACCGTCCAAG  369
          || ||||||||| || |||| |||||||||||||||||||||||||
Db    365 CTGCAGAACAGCCATGGGTTCGGATCTGCCTTCGTCTCTCAGTGCNAG  410

Qy    390 CTGCAGAACAGCCATGGGTTTGGATCTGCCTTCCTCTCAGTGCNAG  435
```

Search completed: April 29, 2003, 09:24:07
Job time : 1924.64 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 16:24:32 ; Search time 139.373 seconds
(without alignments)
9173.565 Million cell updates/sec

Title: US-09-729-264-1
Perfect score: 1175
Sequence: 1 ctgtctgcccatctgaataa...........gtaatacaactgtagtatag 1175

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 709820 seqs, 544064369 residues
Total number of hits satisfying chosen parameters: 1419640

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB ID	Description
1	269	22.9	474	10	US-09-918-995-3342	Sequence 3342, Ap
2	190.4	16.2	401	10	US-09-864-761-16305	Sequence 16305, A
3	142	12.1	398	10	US-09-983-965-4945	Sequence 4945, Ap
4	46	1.9	3577	9	US-10-008-739A-1	Sequence 1, Appli
5	44.8	3.8	1310	9	US-09-849-243-13	Sequence 13, Appl
6	44.8	3.8	2614	9	US-09-822-846-491	Sequence 491, App
7	44.8	3.8	3263	9	US-09-849-243-15	Sequence 15, Appl
8	44.8	3.8	4286	9	US-09-849-243-14	Sequence 14, Appl
9	44.8	3.8	6604	10	US-09-880-107-1748	Sequence 1748, Ap
10	44	9.	455	10	US-09-728-444-151	Sequence 151, App
11	43.4	3.7	384	9	US-09-738-626-544	Sequence 544, App
12	43.4	3.7	3309400	9	US-09-738-626-1	Sequence 1, Appli
13	43.2	3.7	459	10	US-09-864-761-18923	Sequence 18923, A
14	43.2	3.7	459	10	US-09-864-761-2182	Sequence 2182, Ap
15	43.2	3.7	1791	10	US-10-029-217A-2	Sequence 6, Appli
16	42.6	3.6	2424	9	US-09-416-384A-6	Sequence 25, Appli
17	42.6	3.6	4959	9	US-10-029-217A-1	Sequence 1, Appli
18	42.6	3.6	4960	9	US-10-029-217A-30	Sequence 30, Appl
19	42.2	3.6	5959	10	US-09-954-456-1996	Sequence 1996, Ap
20	41.6	3.5	238	10	US-09-864-761-24079	Sequence 24079, A
21	41.6	3.5	554	10	US-09-864-761-7357	Sequence 7357, Ap
22	41.6	3.5	3899	10	US-09-735-367B-5	Sequence 5, Appli
23	41.6	3.5	6504	10	US-09-969-347-171	Sequence 171, App
24	41.6	3.5	6504	10	US-09-735-367B-9	Sequence 9, Appli
25	41.6	3.5	6878	10	US-09-735-367B-1	Sequence 1, Appli
26	41.4	3.5	2572	10	US-09-925-300-486	Sequence 486, App
27	41.2	3.5	424	9	US-09-918-995-10212	Sequence 10212, A
28	41.2	3.5	725	10	US-09-919-580-878	Sequence 878, App
29	41	3.5	2517	10	US-09-476-242-16	Sequence 16, Appl
30	40.8	3.5	2352	10	US-09-476-242-26	Sequence 26, Appl
31	40.8	3.5	2536	9	US-09-918-995-11012	Sequence 11012, A
32	40.8	3.5	6177	10	US-09-917-800A-1546	Sequence 1546, Ap
33	40.8	3.5	6322	10	US-09-918-995-11012	Sequence 11012, A
34	40.6	3.5	547	9	US-09-822-846-129	Sequence 129, App
35	40.4	3.4	954	9	US-10-078-090-83	Sequence 83, Appl
36	40.4	3.4	2674	9	US-09-919-580-11	Sequence 11, Appl
37	40.2	3.4	216	10	US-09-728-444-25	Sequence 25, Appl
38	40.2	3.4	545	10	US-09-919-580-371	Sequence 371, App
39	40.2	3.4	557	10	US-09-919-580-475	Sequence 475, App
40	40.2	3.4	572	10	US-09-919-580-570	Sequence 570, App
41	40.2	3.4	1143	9	US-09-938-842A-1399	Sequence 1399, Ap
42	40.2	3.4	2523	10	US-09-476-242-15	Sequence 15, Appl
43	40.2	3.4	83450	10	US-09-811-469-3	Sequence 3, Appli
44	40	3.4	270	9	US-10-007-557-8	Sequence 8, Appli
45	40	3.4	488	9	US-09-918-995-19918	Sequence 19918, A

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
US-09-918-995-3342

Sequence 3342, Application US/09918995
Publication No. US20030073623A1

GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 3342
LENGTH: 474
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(474)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-3342

Query Match   22.9%;   Score 269;   DB 9;   Length 474;
Best Local Similarity   98.2%;
Matches 272; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy   899   ATTTCAAAAGAAAATCTGAAAAAAGAAGAGAAACANAAAAATCTGAGAGAGAAGTGAAA   958
          |||| |||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   54   ACTTTTAAGGAAAATCTGAAAAAAGAAGAGAAACANAAAAATCTGAGAGAGAAGTGAAA   113

Qy   959   TGAAAAACTCCGGCTACAATTCAGATGAACAAAAGAGACCACCGCTTCTCTCCCTCC   1018
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   114   TGAAAAACTCCGGCTACAATTCAGATGAACAAAAGAGACCACCGCTTCTCTCCCTCC   173

Qy   1019   CAAATCCTGTGAATCCAGTGATCCTGAACAAGAAGAAACAGTAGCTGTGGGCCCTCACCA   1078
          ||||||||||||||||||||||| || |||||||||||||||||||||||||||||||||
Db   174   CAAATCCTGTGAATCCAGTGATCCTGAACAAGAAAGTAGCTGTGTGGCCCTCCTCACCA   233

Qy   1079   GCGGGCTGATCACAGTCCACCCAGGCCAGCAAGTCATCCACAGCGCTTCTTTTAATCTGGC   1138

Qy   234 CAGTCCTCTGAGAAGGTCACTCACTAATACAACTGTAGTAGTAG 293
           ||||||||||||||||||||||||||||||||||||||||||||
Db   1139 CAGTCCTCTGAGAAGGTCACTCACTAATACAACTGTAGTAGTAG 1175

Qy   294 CAGTCCTCTGAGAAGGTCAGTAATACAACTGTAGTATAG 330
           ||||||||||||||||||||||||||||||||||||||||
Db   294 CAGTCCTCTGAGAAGGTCAGTAATACAACTGTAGTATAG 330

RESULT 2
US-09-864-761-16305
; Sequence 16305, Application US20040048763A1
; Patent No. US20040048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 16305
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AF121782.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
US-09-864-761-16305

Query Match   16.2%; Score 190.4; DB 10; Length 401;
Best Local Similarity   97.0%; Pred. No. 2.7e-48;
Matches 194; Conservative   0; Mismatches   6; Indels   0; Gaps   0;

Qy   42 TTCCTCCACGGTTCTGTCTTGGGTTCTGTGGTTAATGAAAGTCATAGAAGAGCCCCCCAGAAATGCAACAGTC 101
           | ||| | ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||| |||
Db   202 TACCTTCCAGGTTCTGTCTTGGGTTCTGTGGTTAATGAAAGTCATAGAAGAGCCCCCAAAATGCAAGAGTC 261

Qy   102 CTGAAGGGCTCCCAGGCGTCGCTTCAACTGCACCGTCTCCCAGGGCTGGGAAGCTCATCATG 161
           | ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   262 CTGAAGGGCTCCCAGGCGTCGCTTCAACTGCACCGTCTCCCAGGGCTGGGAAGCTCATCATG 321

Qy   162 TGGGCTCTCAGTGACATGGTGGTTGCTAAGCGTCAGGCCCATGGAGCCCCATCATCACCAAT 221
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   322 TGGGCTCTCAGTGACATGGTGGTTGCTAAGCGTCAGGCCCATGGAGCCCCATCATCACCAAT 381

Qy   222 GACCGCCTTCACCTCTCAGAG 241
           |||||||||||||||||||||
Db   382 GACCGCCTTCACCTCTCAGAG 401

RESULT 3
US-09-983-965-4945
; Sequence 4945, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 37-21(10297)C
; CURRENT APPLICATION NUMBER: US/09/983,965
; CURRENT FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 09/465,231
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4945
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 26-LIB34-017-Q1-E1-G9
US-09-983-965-4945

Query Match   12.1%; Score 142; DB 10; Length 398;
Best Local Similarity   74.8%; Pred. No. 2.2e-33;
Matches 178; Conservative   0; Mismatches   60; Indels   0; Gaps   0;

Qy   24 GAGATGGGGCTTGTGATTTTCCCCACGGTTCTGGGTCTGGTAATGAAGTCATATAGAAGGC 83
           | ||| ||| ||||||| ||||| ||||||||| ||||| |||||||||||| || |||||
Db   161 GTGCTGGTTGGTGTTCTTGGGCAGCCTGTGGATCCACCAGTGAAATCATAGAGGGT 220

Qy   84 CCCCAGAATGCAACAGTGCCTGAAGGGCTCCCAGGCCTGCGCTTCAACTTGCACCGGCTCTCCCAG 143
           || |||| ||| | |||| ||||| |||||| || | ||||||||||| || ||||
Db   221 CCAAGAAATGTCACAGCCCTGAAGGGGTCGGAGGCTGCGCTTCAACTGCACACTTCTCGGCAG 280

Qy   144 GGCTGGAAGCTCATCATGTGGGCCTCTCAGTGACATGGTGGTTGCTCAGGCCCATG 203
           |||| || ||||| |||||||| ||| ||||| || || ||||| ||||| |||||
Db   281 GGCTGGAAGCTCGTCATGTGGGCCTTCGAGGCACAGTGGTGGCTCAGGCCATGACACCTAAT 340

Qy   204 GAGCCCATCATCACCAATGACCGCTTCACCTCTCAGAGGTACGACAGGGGCGGGAACT 261
           |||||| ||||||||| ||||| |||||||||||| | | ||| |||| || |||
Db   341 GAGACCAATCATCACCAGTGACCGCTTCACTTCGGCCAAGCTACCAACAGAGGGCCGGGAACT 398

RESULT 4
US-10-008-739A-1/c
; Sequence 1, Application US/10008739A
; Patent No. US20020161194A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Inc.
; APPLICANT: Castleberry, Tessa A.
; APPLICANT: Lu, Bihong

APPLICANT: Owen, Thomas A.
APPLICANT: Smock, Steven L.
TITLE OF INVENTION: The Canine Androgen Receptor
FILE REFERENCE: PC10893AGPR
CURRENT APPLICATION NUMBER: US/10/008,739A
CURRENT FILING DATE: 2002-04-15
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin version 3.1

SEQ ID NO 1
LENGTH: 3577
TYPE: DNA
ORGANISM: Canine
US-10-008-739A-1

Query Match   3.9%;   Score 46;   DB 9;   Length 3577;
Best Local Similarity 65.7%;   Pred. No. 0.0036;
Matches 67;   Conservative 0;   Mismatches 35;   Indels 0;   Gaps 0;

Qy   776 GCTTCTGACGCGACGTGTACTCTTACAAATACGCTGGTCGTGCCGCGCCTTCGTTGTTG 835
         ||||||||| |||| ||||||||||||||| |||||||||||||| |||||| ||||||
Db   633 GCTCGTGCTGCTACCTTCTGATACTACCTCCTCGTTGCTGCTGCTGCTGCTGCTGTTGCTG 574

Qy   836 TGGCTGCAACTGCTGCTGCCGGTTGTTGTTTCTGCCTGGTAGAAG 877
         || ||| ||||||||||| |||||||| |||| ||||||||||||
Db   573 TTGCTGCTGCTGCTGCTGCTGCCGCCTGCTGCCGCTGCTGAAGAGGAG 532

RESULT 5
US-09-849-243-13/c
Sequence 13, Application US/09849243
Patent No. US20020157127A1
GENERAL INFORMATION:
APPLICANT: Kirschbaum, Bernd
          Berglund, Erick
          Meisterernst, Michael
          Polites, Greg
TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSCRIPTION
                   COMPLEXES FROM TRANSGENIC
                   NON-HUMAN ANIMALS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: HELLER, EHRMAN, WHITE & McAULIFFE
STREET: 1666 K Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/849,243
FILING DATE: 07-May-2001
ATTORNEY/AGENT INFORMATION:
NAME: Granados, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 38005-0148
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)912-2000
TELEFAX: (202)912-2020
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1310 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: exon
LOCATION: 1..1310
SEQUENCE DESCRIPTION: SEQ ID NO: 13:

US-09-849-243-13

Query Match   3.8%;   Score 44.8;   DB 9;   Length 1310;
Best Local Similarity 81.2%;   Pred. No. 0.0043;
Matches 52;   Conservative 0;   Mismatches 12;   Indels 0;   Gaps 0;

Qy   808 GCTGCTGCTGCTGCCGCCGCCGTCGTTGTTGTTGTGGCTGCAACTGCTGCTGCCGTTGTTTCT 867
         ||||||||||| || | || || ||||| |||| |||| ||||||||||||||||||||
Db   336 GCTGCGCTGCTGCTGTTGCCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGTTGTTGCT 277

Qy   868 GCTG 871
         ||||
Db   276 GCTG 273

RESULT 6
US-09-822-846-491/c
Sequence 491, Application US/09822846
Publication No. US20030021139A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: LaValie, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steininger II, Robert J.
APPLICANT: Bowman, Michael R.
APPLICANT: Spaulding, Vikki
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakar
APPLICANT: Graham, James R.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6400
CURRENT APPLICATION NUMBER: US/09/822,846
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195,605
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 629
SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 491
LENGTH: 2614
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-846-491

Query Match   3.8%;   Score 44.8;   DB 9;   Length 2614;
Best Local Similarity 76.4%;   Pred. No. 0.0068;
Matches 55;   Conservative 0;   Mismatches 17;   Indels 0;   Gaps 0;

Qy   808 GCTGCTGCTGCTGCCGCCGCCGTCGTTGTTGTGGCTGCAACTGCTGCTGCCGTTGTTTTCT 867
         ||| ||||||| ||||| || |||| ||||||||| |||||||||||||||| ||||||||
Db 2179 GCTGTTGGCTGCTGCTGCTGCTGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTTGTTGCTGCT 2120

Qy   868 GCTGTAGAAGAA 879
         ||| ||| || |
Db 2119 GCTGCTGCCGCA 2108

RESULT 7
US-09-849-243-15/c
Sequence 15, Application US/09849243
Patent No. US20020157127A1
GENERAL INFORMATION:
APPLICANT: Kirschbaum, Bernd
          Berglund, Erick

Meisterernst, Michael
Polites, Greg
TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSCRIPTION COMPLEXES FROM TRANSGENIC NON-HUMAN ANIMALS

NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: HELLER, EHRMAN, WHITE & MCAULIFFE
STREET: 1666 K Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/849,243
FILING DATE: 07-May-2001
ATTORNEY/AGENT INFORMATION:
NAME: Granados, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 38005-0148
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)912-2000
TELEFAX: (202)912-2020
FEATURE:
NAME/KEY: exon
LOCATION: 1..3263
SEQUENCE DESCRIPTION: SEQ ID NO: 15:

INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 3263 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

Query Match   3.8%;   Score 44.8;   DB 9;   Length 3263;
Best Local Similarity 81.2%;   Pred. No. 0.0079;
Matches 52; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy   808 GCTGCTGCTGCTGCCGCCGGTTCGTTGTTGTGGCTGCAACTGCTGCCGCCGTTGTTGTTTCT 867
         |||||||||||| |||||||||||||||| || ||||||||||||||||||||| |||| ||
Db 1430 GCTGCTGCTGCTTGCTGCGTTGTTGCTGCTCGCTGCTGCCGCTGCTGTTGTTGTTGTTGCT 1371

Qy   868 GCTG 871
         ||||
Db 1370 GCTG 1367

RESULT 8
US-09-849-243-14/c
Sequence 14, Application US/09849243
Patent No. US20020157127A1
GENERAL INFORMATION:
APPLICANT: Kirschbaum, Bernd
Berglund, Erick
Meisterernst, Michael
Polites, Greg
TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSCRIPTION COMPLEXES FROM TRANSGENIC NON-HUMAN ANIMALS

NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: HELLER, EHRMAN, WHITE & MCAULIFFE
STREET: 1666 K Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA

ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/849,243
FILING DATE: 07-May-2001
ATTORNEY/AGENT INFORMATION:
NAME: Granados, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 38005-0148
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)912-2000
TELEFAX: (202)912-2020
FEATURE:
NAME/KEY: exon
LOCATION: 1..4286
SEQUENCE DESCRIPTION: SEQ ID NO: 14:

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4286 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

Query Match   3.8%;   Score 44.8;   DB 9;   Length 4285;
Best Local Similarity 81.2%;   Pred. No. 0.0095;
Matches 52; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy   808 GCTGCTGCTGCTGCCGCCGGTTCGTTGTTGTGGCTGCAACTGCTGCCGCCGTTGTTGTTTCT 867
         |||||||||||| |||||||||||||||| || ||||||||||||||||||||| |||| ||
Db 2767 GCTGCTGCTGCTTGCTGCGTTGTTGCTGCTCGCTGCTGCCGCTGCTGTTGTTGTTGTTGCT 2708

Qy   868 GCTG 871
         ||||
Db 2707 GCTG 2704

US-09-849-243-14

RESULT 9
US-09-880-107-1748/c
Sequence 1748, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028-WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1748
LENGTH: 6604
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 D83783
US-09-880-107-1748

Query Match   3.8%;   Score 44.8;   DB 10;   Length 6604;
Best Local Similarity 76.4%;   Pred. No. 0.013;
Matches 55; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

RESULT 14
US-10-176-913-584
Sequence 584, Application US/10176913
Publication No. US20030022298A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;   ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C66
; CURRENT APPLICATION NUMBER: US/10/176,913
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-913-584

Query Match   6.1%;   Score 128; DB 9;   Length 708;
Best Local Similarity   23.1%; Pred. No. 0.095;
Matches   53; Conservative 41; Mismatches 95; Indels 40; Gaps 11;

```
Qy   14  SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT  73
             :|  |:|  ||:| |||:|| |:|| ||:| :|||| :|||:||:||||||||
Db   20  AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR-----------DLPG  71

Qy   74  SQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV------QVMGE  122
             |—| :: | | |:| :| || |: |:::| | || |:|          ||:|
Db   72  WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG  131

Qy   123 LFIPSVNLVVAENEPCEVTC-----LPSHWTWLPDISW-ELGLLVSHSSYY--FVPE--  171
             |||:|| | :| ::|          |:| :|| || || :|||| ::||||
Db   132 ---PSVSLVA--GVPANLTCRSRGDARPT-----PELLWFRDGVLLDGATFHQTLLKEGT  181

Qy   172 PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT  220
             :||| ::| |:| :||| || |:||: |:||: :||||:||:||
Db   182 PGSVESTILTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT  230
```

RESULT 15
US-10-180-552-584
Sequence 584, Application US/10180552
Publication No. US20030022300A1
GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE REFERENCE: P3430R1C153
; CURRENT APPLICATION NUMBER: US/10/180,552
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 584
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-180-552-584

Query Match   6.1%;   Score 128; DB 9;   Length 708;
Best Local Similarity   23.1%; Pred. No. 0.095;
Matches   53; Conservative 41; Mismatches 95; Indels 40; Gaps 11;

```
Qy   14  SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT  73
             :|  |:|  ||:| |||:|| |:|| ||:| :|||| :|||:||:||||||||
Db   20  AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR-----------DLPG  71

Qy   74  SQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV------QVMGE  122
             |—| :: | | |:| :| || |: |:::| | || |:|          ||:|
Db   72  WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG  131

Qy   123 LFIPSVNLVVAENEPCEVTC-----LPSHWTWLPDISW-ELGLLVSHSSYY--FVPE--  171
             |||:|| | :| ::|          |:| :|| || || :|||| ::||||
Db   132 ---PSVSLVA--GVPANLTCRSRGDARPT-----PELLWFRDGVLLDGATFHQTLLKEGT  181

Qy   172 PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT  220
             :||| ::| |:| :||| || |:||: |:||: :||||:||:||
Db   182 PGSVESTILTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT  230
```

Search completed: April 29, 2003, 20:17:26
Job time : 47.0659 secs

RESULT 11
US-10-175-752-584
Sequence 584, Application US/10175752
Publication No. US20030022295A1
GENERAL INFORMATION:
;   APPLICANT: Baker, Kevin P.
;   APPLICANT: Chen,Jian
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Pan,James
;   APPLICANT: Smith,Victoria
;   APPLICANT: Watanabe,Colin K.
;   APPLICANT: Wood,William I.
;   APPLICANT: Zhang,Zemin
;   TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;   TITLE OF INVENTION: ACIDS ENCODING THE SAME
;   FILE REFERENCE: P3430R1C60
;   CURRENT APPLICATION NUMBER: US/10/175,752
;   CURRENT FILING DATE: 2002-06-19
;   Prior Application removed - See File Wrapper or Palm
;   NUMBER OF SEQ ID NOS: 612
;   SEQ ID NO 584
;   LENGTH: 708
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-175-752-584

Query Match   6.1%;   Score 128;   DB 9;   Length 708;
Best Local Similarity   23.1%;   Pred. No. 0.095;
Matches 53; Conservative 41; Mismatches 95; Indels 40; Gaps 11;

Qy   14 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
            :| :::| ||| :| ||||| |||| |::: :|| : | : || :|: :| :|||:
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR------------ 71

Qy   74 SQRYDQGGNFTS----EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV---------QVMGE 122
         || | ||| | | : || | | ||||: |||||| | |:| ||:| | | |
Db   72 WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy   123 LFIPSVNLVVAENEPCEVTC------LPSHWTWLPDISW-ELGLLVSHSSSY--FVPE-- 171
           |||:|| | |:::| :| |:| || |||| | : || | | |:
Db   132 ---PSVSLVA--GVPANLTCRSRGDRAPT-----PELLWFRDGVLLDGATFHQTLLKEGT 181

Qy   172 PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 220
              || : || |: |: :||| | : ::| |: |: || | |:
Db   182 PGSVESTLILTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVI 230

RESULT 12
US-10-176-482-584
Sequence 584, Application US/10176482
Publication No. US20030022296A1
GENERAL INFORMATION:
;   APPLICANT: Baker, Kevin P.
;   APPLICANT: Chen,Jian
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Pan,James
;   APPLICANT: Smith,Victoria
;   APPLICANT: Watanabe,Colin K.
;   APPLICANT: Wood,William I.
;   APPLICANT: Zhang,Zemin
;   TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;   TITLE OF INVENTION: ACIDS ENCODING THE SAME
;   FILE REFERENCE: P3430R1C70
;   CURRENT APPLICATION NUMBER: US/10/176,482
;   CURRENT FILING DATE: 2002-06-20
;   Prior Application removed - See File Wrapper or Palm

RESULT 13
US-10-176-757-584
Sequence 584, Application US/10176757
Publication No. US20030022297A1
GENERAL INFORMATION:
;   APPLICANT: Baker, Kevin P.
;   APPLICANT: Chen,Jian
;   APPLICANT: Desnoyers,Luc
;   APPLICANT: Goddard,Audrey
;   APPLICANT: Godowski,Paul J.
;   APPLICANT: Gurney,Austin L.
;   APPLICANT: Pan,James
;   APPLICANT: Smith,Victoria
;   APPLICANT: Watanabe,Colin K.
;   APPLICANT: Wood,William I.
;   APPLICANT: Zhang,Zemin
;   TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;   TITLE OF INVENTION: ACIDS ENCODING THE SAME
;   FILE REFERENCE: P3430R1C86
;   CURRENT APPLICATION NUMBER: US/10/176,757
;   CURRENT FILING DATE: 2002-06-20
;   Prior Application removed - See File Wrapper or Palm
;   NUMBER OF SEQ ID NOS: 612
;   SEQ ID NO 584
;   LENGTH: 708
;   TYPE: PRT
;   ORGANISM: Homo Sapien
US-10-176-757-584

Query Match   6.1%;   Score 128;   DB 9;   Length 708;
Best Local Similarity   23.1%;   Pred. No. 0.095;
Matches 53; Conservative 41; Mismatches 95; Indels 40; Gaps 11;

Qy   14 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
            :| :::| ||| :| ||||| |||| |::: :|| : | : || :|: :| :|||:
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGLALGGQR------------ 71

Qy   74 SQRYDQGGNFTS----EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV---------QVMGE 122
         || | ||| | | : || | | ||||: |||||| | |:| ||:| | | |
Db   72 WSRYWISGNAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy   123 LFIPSVNLVVAENEPCEVTC------LPSHWTWLPDISW-ELGLLVSHSSSY--FVPE-- 171
           |||:|| | |:::| :| |:| || |||| | : || | | |:
Db   132 ---PSVSLVA--GVPANLTCRSRGDRAPT-----PELLWFRDGVLLDGATFHQTLLKEGT 181

Qy   172 PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 220
              || : || |: |: :||| | : ::| |: |: || | |:
Db   182 PGSVESTLILTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVI 230

Publication No. US20030131S3A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Chen,Jian
APPLICANT: Desnoyers,Luc
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C50
CURRENT APPLICATION NUMBER: US/10/175,737
CURRENT FILING DATE: 2002-06-19
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
LENGTH: 708
TYPE: PRT
ORGANISM: Homo Sapien
US-10-175-737-584

Query Match   6.18;   Score 128; DB 9;   Length 708;
Best Local Similarity   23.1%; Pred. No. 0.095;
Matches 53; Conservative 41; Mismatches 95; Indels 40; Gaps 11;

```
Qy  14 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
          :| ||||||  | ||:|  ||||: ||||||||||||||||||||||||||||||
Db  20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGIALGGQR------DLPG 71

Qy  74 SQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV------QVMGE 122
         ||| |:|    | | :||:|  |||  |:|  |:| ||  : ||      |||
Db  72 WSRYWISGNAANQCHDLHIRPVELEDEASYECQATOAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy 123 LFIPSVNLVVAENEPCEVTC-----LPSHWTWLPDISW-ELGLLVSHSSYY--FVPE-- 171
        ||:|  |    |: |  |      |:    |  | ||:     ||:|  ::: ||
Db 132 ---PSVSLVA--GVPANLTCRSRGDARPT-----PELLWFRDGVLLDGATFHQTLLKEGT 181

Qy 172 PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 220
         ||| ||  |  | |     |  |    | | |    |   | :|| |
Db 182 PGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 9
US-10-173-706-584
Sequence 584, Application US/10173706
Publication No. US20030022293A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Chen,Jian
APPLICANT: Desnoyers,Luc
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C7
CURRENT APPLICATION NUMBER: US/10/173,706
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612

; ORGANISM: Homo Sapien
US-10-173-706-584

RESULT 10
US-10-175-738-584
; Sequence 584, Application US/10175738
; Publication No. US20030022294A1
GENERAL INFORMATION:
APPLICANT: Baker,Kevin P.
APPLICANT: Chen,Jian
APPLICANT: Desnoyers,Luc
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Smith,Victoria
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C45
CURRENT APPLICATION NUMBER: US/10/175,738
CURRENT FILING DATE: 2002-06-19
Prior application.removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
LENGTH: 708
TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-738-584

Query Match   6.18;   Score 128; DB 9;   Length 708;
Best Local Similarity   23.1%; Pred. No. 0.095;
Matches 53; Conservative 41; Mismatches 95; Indels 40; Gaps 11;

```
Qy  14 SGSGNEVIEGPQNARVLKGSQARFNCTVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
          :| ||||||  | ||:|  ||||: ||||||||||||||||||||||||||||||
Db  20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWGLVQWTKSGIALGGQR------DLPG 71

Qy  74 SQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV------QVMGE 122
         ||| |:|    | | :||:|  |||  |:|  |:| ||  : ||      |||
Db  72 WSRYWISGNAANQCHDLHIRPVELEDEASYECQATOAGLRSRPAQLHVLVPPEAPQVLGG 131

Qy 123 LFIPSVNLVVAENEPCEVTC-----LPSHWTWLPDISW-ELGLLVSHSSYY--FVPE-- 171
        ||:|  |    |: |  |      |:    |  | ||:     ||:|  ::: ||
Db 132 ---PSVSLVA--GVPANLTCRSRGDARPT-----PELLWFRDGVLLDGATFHQTLLKEGT 181

Qy 172 PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 220
         ||| ||  |  | |     |  |    | | |    |   | :|| |
Db 182 PGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

```
Db   483  SSSFSVS 489
          :||:||:
```

RESULT 5

US-09-817-487A-2
Sequence 2, Application US/09817487A
Patent No. US20020150876A1
GENERAL INFORMATION:
  APPLICANT: No. US20020150876A1artis AG
  TITLE OF INVENTION: Selectable Marker Genes
  FILE REFERENCE: 4-31193A
  CURRENT APPLICATION NUMBER: US/09/817,487A
  CURRENT FILING DATE: 2002-02-14
  NUMBER OF SEQ ID NOS: 10
  SOFTWARE: Patentin version 3.0
SEQ ID NO 2
  LENGTH: 869
  TYPE: PRT
  ORGANISM: homo sapiens
US-09-817-487A-2

Query Match
Best Local Similarity 24.3%;   Score 131.5; DB 10; Length 869;
Matches 57; Conservative 36; Mismatches 95; Indels 47; Gaps 11;

```
Qy    4 GAMENRDPPGSGSGNEVIEGPQNARVLKGSQARFNCIVSQGWK-LIMWALSDMVVLSVRP 62
Db  113 GALQVKMKP-------KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD------- 158

Qy   63 MEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMG 121
Db  159 -SPLRENSRIAVLE------SGSLRIHNVQKEDAGQYRCVAKNSL--GTAVSKVVKLEF 208

Qy  122 ELFIPSVNLVVAENEP----------CEVTCLPSHWTWLPDISW-ELGLLVSHSSYYFVPEP 172
Db  209 EVFARILRAPESHNVTFGSSFVTLHCTATGIP-------VPTITWIENGNAVSSGSIQESVKD 263

Qy  173 SDLQSAVSILALTPQSNGTLTCVAT------WKSLKARKSATVNLTVIRCPQDTGG 223
Db  264 RVIDSRLQLFITKP---GLYTCIATNKHGEKFSTAKAAATISIAEWSKPQKDNKG 315
```

RESULT 6

US-10-174-590-584
Sequence 584, Application US/10174590
Publication No. US20030008352A1
GENERAL INFORMATION:
  APPLICANT: Baker,Kevin P.
  APPLICANT: Chen,Jian
  APPLICANT: Desnoyers,Luc
  APPLICANT: Goddard,Audrey
  APPLICANT: Godowski,Paul J.
  APPLICANT: Gurney,Austin L.
  APPLICANT: Pan,James
  APPLICANT: Smith,Victoria
  APPLICANT: Watanabe,Colin K.
  APPLICANT: Wood,William I.
  APPLICANT: Zhang,Zemin
  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
    ACIDS ENCODING THE SAME
  FILE REFERENCE: P3430R1C42
  CURRENT APPLICATION NUMBER: US/10/174,590
  CURRENT FILING DATE: 2002-06-18
  Prior application removed - See File Wrapper or Palm
  NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
  LENGTH: 708
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-174-590-584

Query Match   6.1%;   Score 128; DB 9; Length 708;

RESULT 7

US-10-176-758-584
Sequence 584, Application US/10176758
Publication No. US20030008353A1
GENERAL INFORMATION:
  APPLICANT: Baker,Kevin P.
  APPLICANT: Chen,Jian
  APPLICANT: Desnoyers,Luc
  APPLICANT: Goddard,Audrey
  APPLICANT: Godowski,Paul J.
  APPLICANT: Gurney,Austin L.
  APPLICANT: Pan,James
  APPLICANT: Smith,Victoria
  APPLICANT: Watanabe,Colin K.
  APPLICANT: Wood,William I.
  APPLICANT: Zhang,Zemin
  TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
    ACIDS ENCODING THE SAME
  FILE REFERENCE: P3430R1C104
  CURRENT APPLICATION NUMBER: US/10/176,758
  CURRENT FILING DATE: 2002-06-21
  Prior Application removed - See File Wrapper or Palm
  NUMBER OF SEQ ID NOS: 612
SEQ ID NO 584
  LENGTH: 708
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-176-758-584

Query Match
Best Local Similarity 23.1%;   Pred. No. 0.095;
Matches 53; Conservative 41; Mismatches 95; Indels 40; Gaps 11;

```
Qy   14 SGSGNEVIEGPQNARVLKGSQARFNCIVSQGWKLIMWALSDMVVLSVRPMEPIITNDRFT 73
Db   20 AGPSPHFLQQPEDLVVLLGEEARLPCALGAYWLVQWTKSGLALGGQR------DLPG 71

Qy   74 SQRYDQGGNFTS---EMIIHNVEPSDSGNIRCSLQNSRLHG-SAYLTV------QVMGE 122
Db   72 WSRYWISGNAAANGQHDLHIRPVELEDEASYECQATQAGLRSRPAQLHV----PPEAPQVLGG 131

Qy  123 LFIPSVNLVVAENEPCEVTC-----LPSHWTWLPDISW-ELGLLVSHSSYY--FVPE-- 171
Db  132 ---PSVSLVA--GVPANLTCRSRGDARPT-----PELLWFRDGVLLDGATFHQTLLKEGT 181

Qy  172 PSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDT 220
Db  182 PGSVESTLTLTPFSHDDGATFVCRARSQALPTGRDTAITLSLQYPPEVT 230
```

RESULT 8

US-10-175-737-584
Sequence 584, Application US/10175737

Db   362 ILMQRR------------------QRRGEERKAPENQEEEERAELN-------- 390

Qy   325 QKTTDTASLPPKSCESSDPEQRNSSCGPP 353

Db   391 ----------QSEEPEAGSSSTGP 405

RESULT 2
US-10-184-644-559
; Sequence 559, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 559
; LENGTH: 2473
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-559

  Query Match   6.6%;   Score 138.5; DB 9; Length 2473;
  Best Local Similarity 29.5%; Pred. No. 0.069;
  Matches 31; Conservative 3; Mismatches 36; Indels 35; Gaps 2;

Qy   183 ALTPQSNGTLTCVATWKSLKARKSATVNLTIVIRCPQDTGGGINIPGVLSSLPSLGFSLPT 242
              | | | || |||||||||||||||||||||||||| ||| |||| ||||||||
Db   2274 AATTGAAGTTTCAATTAAAATTTAATATGTTTCC---------------------------- 2307

Qy   243 WGKVGLGLAGTMLLTPTCTLTIRCCCCR-RRCCGCNCCCRCCFCC 286
              |||| ||| ||||| |||| ||| | | | | | |||
Db   2308 ------ATTCCTCATCGCCCACCCCCCGGCCCCCCCCCCACCACC 2344

RESULT 3
US-10-184-634-559
; Sequence 559, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634

; LENGTH: 2473
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-559

  Query Match   6.6%;   Score 138.5; DB 9; Length 2473;
  Best Local Similarity 29.5%; Pred. No. 0.069;
  Matches 31; Conservative 3; Mismatches 36; Indels 35; Gaps 2;

Qy   183 ALTPQSNGTLTCVATWKSLKARKSATVNLTIVIRCPQDTGGGINIPGVLSSLPSLGFSLPT 242
              | | | || |||||||||||||||||||||||||| ||| |||| ||||||||
Db   2274 AATTGAAGTTTCAATTAAAATTTAATATGTTTCC---------------------------- 2307

Qy   243 WGKVGLGLAGTMLLTPTCTLTIRCCCCR-RRCCGCNCCCRCCFCC 286
              |||| ||| ||||| |||| ||| | | | | | |||
Db   2308 ------ATTCCTCATCGCCCACCCCCCGGCCCCCCCCCCACCACC 2344

RESULT 4
US-10-016-283-33
; Sequence 33, Application US/10016283
; Patent No. US20020164702A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al., David M.
; TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTORS AND LIGANDS
; FILE REFERENCE: REG195-B-PCT-US
; CURRENT APPLICATION NUMBER: US/10/016,283
; CURRENT FILING DATE: 2001-11-30
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: US/09/077,955A
; PRIOR FILING DATE: 1996-12-13
; PRIOR APPLICATION NUMBER: PCT/US96/20696
; PRIOR FILING DATE: 1996-11-14
; NUMBER OF SEQ ID NOS: 36
; SEQ ID NO 33
; LENGTH: 869
; TYPE: PRT
; ORGANISM: Homo sapiens
; SOFTWARE: PatentIn Ver. 2.0
US-10-016-283-33

  Query Match   6.3%;   Score 132.5; DB 9; Length 869;
  Best Local Similarity 20.8%; Pred. No. 0.056;
  Matches 89; Conservative 55; Mismatches 178; Indels 105; Gaps 20;

Qy   4 GAMENRDPPGSGSSGNEVIEGPQNARVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRP 62
             : || :| | | | | ::|| : ||| || | |: |: || :| | |
Db   113 GALQVKMKP-------KITRPPINVKIIEGLRAVLPCTTMGNPKPSVSWIKGD------- 158

Qy   63 MEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMG 121
             |: | || || | |:| : |||::|||:|::||:|||| |||: ||::|
Db   159 -SPLRENSRIAVLE------SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKVVKLEV 208

Qy   122 ELFIPSVNLVVAENEP------------CEVTCLPSHWTWLPDISW-ELGLLVSHSSYFVFPEP 172
             ||: :: ||:| | ||:|| ||: |||| || :| |: :| | ::| ||
Db   209 EVFARILRAPESHNVFGSFVTLHCTATGIP----------VPTIWIENGNAVSGSGSIQESVKD 263

Qy   173 SDLQSAVSILALTPQSNGTLTCVAT----WKSLKARKSATVNLTIVIRCPQDTGGG----- 223
             || | | :| || | ||||||:| | |||||||||||||||||||||||||
Db   264 RVIDSRLQLFITKP---GLYTCIATNKHGEKFSTAKAAATISIAEWSKPQKDNKGYCAQY 320

Qy   224 ----INIPGVLSSLPSLGFSL----------------PTWGKVGLGLAGTMLLTPTCTLTIRCC 267
                    | |||:||| |||| ::| |
Db   321 RGEVCNAVLAKDALVFLNTSYADPEEAQELIVHTAWNEL------KVVSPVCRPAAEAL 373

Qy   268 CCRRRCGCN------------CCCRCC-----FCCRRKRGRFRIFQFOKKSEKEKTNKETETE 312
             | : :| :| |: :|| :||:|
Db   374 LCNHIFQECSPGVVPFPIPICREYCLAVKELFCAKE---WLVMEEKTHRGLYRSEMHLL 429

Qy   313 SGNENGGYNSDEQKTTDTASLPPKSCESSDPEQRNSSCGPPHQRADQRP----PRPASHP 368
             || : | | | || | | | :|:|| | |:| : | | ||:|
Db   430 SVPECSKLPSHMWDPPACARLP--------HLDYNKENLKTFPP--MTSSKPSVDIPNLPSSS 482

Qy   369 QASFNLA 375

OM protein - protein search, using sw model

   GenCore version 5.1.4_p5_4578
   Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 29, 2003, 11:05:56 ; Search time 19.0659 seconds
           (without alignments)
           1622.277 Million cell updates/sec

Title: US-09-729-264-4
Perfect score: 2088
Sequence: 1 MVAGAMENRDPPGSGSGNEV...........HPQASFNLASPEKVSNPTVV 386

Scoring table: BLOSUM62
      Gapop 10.0 , Gapext 0.5
Searched: 301932 seqs, 80129803 residues

Total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
      Maximum Match 100%
      Listing first 45 summaries

Database : Published_Applications_AA:*
  1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
  2: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
  3: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
  4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
  5: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
  6: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
  7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
  8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
  9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
  10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
  11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
  12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
  13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
  14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match Length	DB ID	Description
1	140	405 1	US-08-755-235-4	Sequence 4, Appli
2	138.5	6.7 2473 9	US-10-184-644-559	Sequence 559, App
3	138.5	6.6 2473 9	US-10-184-634-559	Sequence 559, App
4	132.5	6.3 869 9	US-10-016-283-33	Sequence 33, Appl
5	131.5	6.3 869 10	US-09-817-487A-2	Sequence 2, Appli
6	128	6.1 708 9	US-10-174-590-584	Sequence 584, App
7	128	6.1 708 9	US-10-176-758-584	Sequence 584, App
8	128	6.1 708 9	US-10-175-738-584	Sequence 584, App
9	128	6.1 708 9	US-10-173-706-584	Sequence 584, App
10	128	6.1 708 9	US-10-175-737-584	Sequence 584, App
11	128	6.1 708 9	US-10-176-757-584	Sequence 584, App
12	128	6.1 708 9	US-10-176-482-584	Sequence 584, App
13	128	6.1 708 9	US-10-176-752-584	Sequence 584, App
14	128	6.1 708 9	US-10-180-552-584	Sequence 584, App
15	128	6.1 708 9	US-10-176-913-584	Sequence 584, App
16	128	6.1 708 9	US-10-180-557-584	Sequence 584, App
17	128	6.1 708 9	US-10-173-700-584	Sequence 584, App
18	128	6.1 708 9	US-10-174-572-584	Sequence 584, App
19	128	6.1 708 9	US-10-174-579-584	Sequence 584, App
20	128	6.1 708 9	US-10-174-582-584	Sequence 584, App
21	128	6.1 708 9	US-10-174-588-584	Sequence 584, App
22	128	6.1 708 9	US-10-175-739-584	Sequence 584, App
23	128	6.1 708 9	US-10-175-740-584	Sequence 584, App
24	128	6.1 708 9	US-10-175-743-584	Sequence 584, App
25	128	6.1 708 9	US-10-176-488-584	Sequence 584, App
26	128	6.1 708 9	US-10-176-492-584	Sequence 584, App
27	128	6.1 708 9	US-10-176-747-584	Sequence 584, App
28	128	6.1 708 9	US-10-176-750-584	Sequence 584, App
29	128	6.1 708 9	US-10-176-985-584	Sequence 584, App
30	128	6.1 708 9	US-10-176-987-584	Sequence 584, App
31	128	6.1 708 9	US-10-176-991-584	Sequence 584, App
32	128	6.1 708 9	US-10-176-992-584	Sequence 584, App
33	128	6.1 708 9	US-10-176-993-584	Sequence 584, App
34	128	6.1 708 9	US-10-184-658-584	Sequence 584, App
35	128	6.1 708 9	US-10-173-695-584	Sequence 584, App
36	128	6.1 708 9	US-10-173-697-584	Sequence 584, App
37	128	6.1 708 9	US-10-173-705-584	Sequence 584, App
38	128	6.1 708 9	US-10-174-576-584	Sequence 584, App
39	128	6.1 708 9	US-10-174-585-584	Sequence 584, App
40	128	6.1 708 9	US-10-174-586-584	Sequence 584, App
41	128	6.1 708 9	US-10-175-747-584	Sequence 584, App
42	128	6.1 708 9	US-10-176-481-584	Sequence 584, App
43	128	6.1 708 9	US-10-176-485-584	Sequence 584, App
44	128	6.1 708 9	US-10-176-487-584	Sequence 584, App
45	128	6.1 708 9	US-10-176-493-584	Sequence 584, App

ALIGNMENTS

RESULT 1
US-08-755-235-4

Sequence 4, Application US/08755235
; Publication No. US20030059423A1
; GENERAL INFORMATION:
; APPLICANT: Wu, Jun
; APPLICANT: Stern, David M.
; APPLICANT: Schmidt, Ann Marie
; TITLE OF INVENTION: METHOD FOR TREATING SYMPTOMS OF DIABETES
; FILE REFERENCE: 0575/50159
; CURRENT APPLICATION NUMBER: US/08/755,235
; CURRENT FILING DATE: 1996-11-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patentin version 3.1
;
; SEQ ID NO 4
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Human
US-08-755-235-4

 Query Match   6.7%; Score 140; DB 1; Length 405;
 Best Local Similarity 23.7%; Conservative 34; Pred. No. 0.006;
 Matches 78; Mismatches 93; Indels 124; Gaps 16;

Qy 64 EPIITNDRFTS-----QRYDQGGNFT--SEMIIHNVEPSDSGNIR----CSLQNSRLHGS 112
     :||::|| :||:| :||:| |||||| ||| ||:| ||| |:|
Db 162 KPLVPNEKGVSVKEQTRRHPETGLFTLQSELM--VTPARGGDRPTFSCSFSPGLRHR 218

Qy 113 AYLTVQVMGELFIP-----SVNLVVAENEP-----------CEVTCLPSHWTWLPDIS 154
     ||| | | ||| | || || || | | || | | :||
Db 219 ALRTAPIQPRVWEEPVPLEEVQLVV--EPEGGAVAPGGTVTLLTCEVPAQPS----PQIH 270

Qy 155 WELGLLVSHSSSYYFVPEPSDLQSAVSILALTPOSNGTLTCVATWKSLRARKS-ATVNLTV 213
     | ||| || |||| : | | | || | | ||| ||
Db 271 WMKD-------GVPLPLPPSPVLILPEIGPQDGTYSCVATHSSHGPQESRAVVSISI 321

Qy 214 IRCPDTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLLT------PTCTLIIRC 266
      ||| | ||| || | :| || |:|:: |||
Db 322 IE-PGEEG---------PTAGSVGGSGLGTLALALGILGGLGTAALLGV 361

Qy 267 CCCRRRCCGNCCCRCCFCCRRKRGFRIQPKKSEKEKT--NKETETESGNENSGYNSDE 324
     :|| :|:| | | |:| | |

```
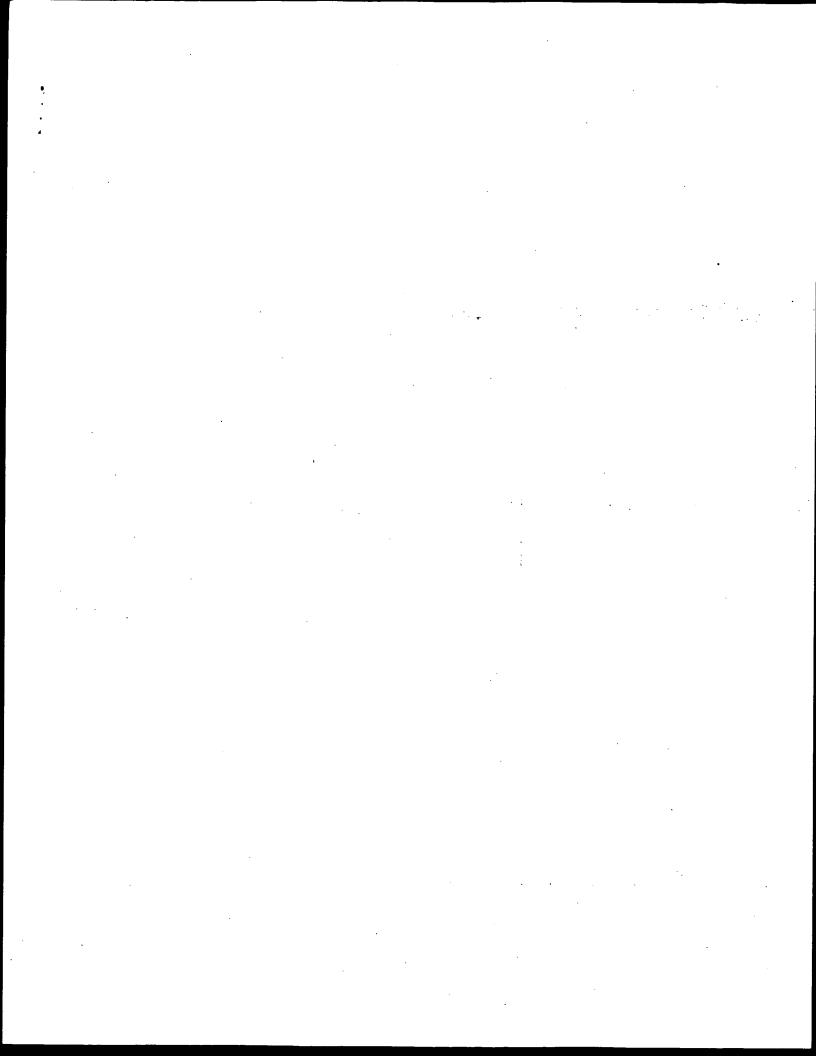C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; prc
C;Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyca
structural protein
F;24752-25008/Domain: protein kinase homology <KIN>
F;84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,4068,
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1
tatus predicted
F;16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680,18
,21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248
F;26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match         5.5%;  Score 114.5;  DB 1;  Length 26926;
Best Local Similarity  22.1%;  Pred. No. 45;
Matches    62;  Conservative  31;  Mismatches  87;  Indels  101;  Gaps   12;

Qy      3  AGAMENRDPPGSGSGGNEVIEGPQ-----------------------NARV------------LKGSQARFNCTVS 42
           ||  |: | | |                                    || |              | ||
Db  16422  AGVGEHADVPGPIIVEEKLEAPDIDLDLELRKIINIRAGGSLRLFVPIKGRP-----TPE 16476

Qy     43  QGWKLIMWALSDMVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC 102
           ||  || : :|                       |||| ::: | || |||||||||||
Db  16477  VKWGKVDGEIRDAAIIDVI-----------------SSFTS-LVLDNVNRYDSGKYTL 16516

Qy    103  SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTWLPDI---------- 153
           :|:||   ||::||:|: |||||                         ::
Db  16517  TLENSSGTKSAFVTVRVL-DTPSPPVNLKVTEITKDSVSI----TWEPPLLDGGSKIKN 16570

Qy    154  -----------------------SWELGLLVSHSSYYF--------------VPEPSDLQSAVSI 181
                                   ||                          ::
Db  16571  YIVEKREATRKSYAAVTNCHKNSWKIDQLQEGCSYYFRVTAENEYGIGLPAQTADPIKV 16630

Qy    182  LALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTCG 222
           | |: |    | | ::  | ||              |
Db  16631  -AEVPQPPGKITVDDV------TRNSVSLSWTK-----PEHDGG 16662

Search completed: April 28, 2003, 21:09:09
Job time : 40.0659 secs
```

RESULT 14.
I49583

differentiation antigen - mouse
C;Species: Mus musculus (house mouse)
A;Title: Organization of the murine Cd22 locus. Mapping to chromosome 7 and characteriza[tion]
A;Reference number: I49583; MUID:93315834; PMID:8100843
A;Accession: I49583
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-862 <RES>
A;Cross-references: GB:L16928; NID:g348965; PIDN:AAA02562.1; PID:g348966
C;Genetics:
A;Gene: CD22

J. Immunol. 151, 175-187, 1993
R;Law, C.

```
Db   145  TPEVKWGKVDGDIRDAAIIDVTSSFTS-LVLDNVNRYDSGKYTLTLENSSGTKSAFVTVR  203
Qy   119  VMGELFIPSVNLVVAENEPCEVTCLPSHWTWLPDI-------------------------  153
Db   204  VL-DTPSPPVNLKVTEITKDSVSI-----TWEPPLLDGGSKIKNYIVEKREATRKSYAAV  257
Qy   154  -----SWELGLLVSHSSYYFVPEPSDLQSAVSILALT----------PQSNGTLTCVAT  197
Db   258  VTNCHKNSWKIDQLOEGCSYYF-RVIAENEYGIGLAARTADDPIKVAEVPQPPGKITVVDDV  316
Qy   198  WKSLKARKSATVNLTVIRCPQDTGG  222
Db   317  -----TRNSVSLSWTK--PEHDGG  333
```

Query Match   5.5%;
Best Local Similarity   19.1%;
Matches 86; Conservative 36;
                             Score 114.5; DB 2; Length 862;
                             Pred. No. 1.1;
                             Mismatches 126; Indels 203; Gaps 18;

```
Qy    10  DPPGSGSGNEVIEGPQNARVLKGSQAREN----------CTVSQGWKL----IMWA   51
Db   472  NPQGSGS----------VLKPGVLRIQKVTWDSMPVSCAACNHKCSWALPVILNVHYA  519

Qy    52  LSDMVVLSVRPMEPIIINDRFTSQ-----------RYDQGNFTSE--MIIHNVEPS   95
Db   520  PRDVKVLKVSPASEIRAQRVLLQCDFAESNPAEVRFFWKKNGSLVQEGRYLSFGSVSPE  579

Qy    96  DSGNIRCSLQNSRLHGSAYLTVQVMGELFIPSVNLV-----------VAENEPC  138
Db   580  DSGNYNCMVNNS-----------IGETLSQAWNLQVLYAPRRLRVSISPGDHVMEGKKA  627

Qy   139  EVTCLP------SHWTWLPDISWELGLLVSHSSYYFVPEPSDLQSAVSILALTP---QSN  189
Db   628  TLSCESDANPPISQYTW-----------FDSSGQDLHSSGQKLRLEPLEVQHT  669

Qy   190  GTLTCVATWKSLKARKSATVNLTVIRCPQDTGGINIPGVLSSLPSLGFSLPTWGKVGLG  249
Db   670  GSYRCKGT-NGIGTGESPPSTLIVYSPETIG---------KRVALG  706

Qy   250  LAGTMLLPPTCLTLTIRCCCCRRRCGCNCCCRCCFCCRRKRGFRIQ----FQK  298
Db   707  LGFCLTI---CILAI--------------WGMKIQKKWKQNRSQQGLQE  738

Qy   299  KSE------KEKTNKEITEISGNENSG-----YNSDEQKTTDT  330
Db   739  NSSGQSFFVYNKKARTPLSEGPQSQGYNPAMDDTVSYAILRFPESDMHNAGDAGTPAT  798

Qy   331  ASLPPKSCES------------SDPEQRNSSC  350
Db   799  QAPPNNSDSVTYSVIQKRPMGDYENVNPSC  829
```

RESULT 15
I38344
titin, cardiac muscle [validated] - human

N;Alternate names: connectin
N;Contains: serine/threonine-specific protein kinase (EC 2.7.1.-)
C;Species: Homo sapiens (man)
C;Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 15-Sep-2000
A;Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393
R;Labeit, S.; Kolmerer, B.
A;Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
Science 270, 293-296, 1995
A;Reference number: A57430; MUID:96026330; PMID:7559978
A;Accession: I38344
A;Status: nucleic acid sequence not shown; translation not shown; translated from GB/
A;Molecule type: mRNA
A;Residues: 1-26926 <LAB1>
A;Cross-references: EMBL:X90568; NID:g1017424; PID:g1017425
R;Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A;Title: Dissecting titin into its structural motifs: identification of an alpha-heli
A;Reference number: I38345; MUID:95119041; PMID:7819249
A;Accession: I38345
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1977-2014 <MUS>
A;Cross-references: EMBL:X83270; NID:g602579; PIDN:CAA58243.1; PID:g602580
A;Note: conformation and properties are reported for a synthetic peptide correspondin
R;Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A;Title: Towards a molecular understanding of titin.
A;Reference number: S20897; MUID:92258380; PMID:1582406
A;Accession: S20898
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 13597-14200,'I',14202-14696 <LAB2>
A;Cross-references: EMBL:X64698; NID:g937192; PIDN:CAA45939.1; PID:g937193
A;Accession: S20897
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 16330-16382,'S',16384-16756,'F',16758-16860 <LAB3>
A;Cross-references: EMBL:X64699; NID:g937190; PIDN:CAA45940.1; PID:g937191
A;Accession: S20899
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: mRNA
A;Residues: 'P',22278-22431,'R',22433-22448,'Q',22450-22453,'Q',22455-22480,'TR',2248
A;Cross-references: EMBL:X64697; NID:g937190; PIDN:CAA45938.1; PID:g937195
R;Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.
J. Mol. Biol. 256, 556-563, 1996
A;Title: Genomic organization of M line titin and its tissue-specific expression in t
A;Reference number: S63665; MUID:96177761; PMID:8604138
A;Accession: S63665
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 26729-26825 <KOL>
A;Cross-references: EMBL:X92412; NID:g1236761
R;Gautel, M.; Leonard, K.; Labeit, S.
EMBO J. 12, 3827-3834, 1993
A;Title: Phosphorylation of KSP motifs in the C-terminal region of titin in different
A;Reference number: S37393; MUID:94008990; PMID:8404852
A;Accession: S37393
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 26831-26926 <GAU>
A;Cross-references: EMBL:X92412; NID:g1236761
R;Improta, S.; Politou, A.S.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, February 1996
A;Reference number: A66736; PDB:1TIT
A;Contents: annotation; conformation by (1)H-NMR, residues 5253-5341
R;Pfuhl, M.; Pastore, A.
submitted to the Brookhaven Protein Data Bank, August 1996
A;Reference number: A62201; PDB:1NCT
A;Contents: annotation; conformation by (1)H-NMR, residues 'S',26059-26155
C;Genetics:
A;Gene: GDB:TTN
A;Cross-references: GDB:127867; OMIM:188840
A;Map position: 2q31-2q32
C;Function:
A;Description: structural protein forming filaments in striated muscle

C;Keywords: cell adhesion; extracellular protein; glycoprotein; phosphotransferase; tran
F;1-30/Domain: signal sequence #status predicted <SIG>
F;31-1070/Product: protein-tyrosine kinase 7 #status predicted
F;31-703/Domain: extracellular #status predicted <EXC>
F;704-725/Domain: transmembrane #status predicted <TMM>
F;704-1070/Domain: intracellular #status predicted <INC>
F;726-1070/Domain: protein kinase homology <KIN>
F;794-1065/Domain: protein kinase homology <KIN>
F;116,175,184,214,268,283,405,463,567,646/Binding site: carbohydrate (Asn) (covalent) #s

Query Match   5.7%;   Score 119; DB 2; Length 1070;
Best Local Similarity 25.3%; Pred. No. 0.66;
Matches 58; Conservative 32; Mismatches 103; Indels 36; Gaps 10;

```
Qy   20 VIEGPQNARVLKGSQOARFNCTVS-QGWKLIMWALSDMVVLSVRPMEPTITNDRFTSQRYD  78
          :|| .| |:|.:|| | :||.| :|||:| |||||||||| ||||  |||||||||||||
Db  227 VVLAPQDVVVARYEEAMFHCQFSAQPPSLQWLFEDETPITNRSRPHLRRATVFA----   282

Qy   79 QGGNFTSEMIIHNVEPSDSGNIRCSLQNSR------LHGSAYLTVQVMGELFIPSVNLLVA 133
           ||: .:   |     .|| | :  : :         |||| |||| |
Db  283 -----NGSLLTQVRPRNAGIYRCIGQOGRGPPIILEATHLAEIEDMPLFEPRVFTAGS   337

Qy  134 ENEPCEVTCLPSHWTWLPDISWE-LGL-LVSHSSYYFVPEPSDLQSAVSILALTPQSN-G 190
          | |: |::|| |    |    .|  |:| ||||||||||||||
Db  338 EE--RVTCLPPKGLPEPSVWWEHAGVRLPTHGRVY-----QKGHELVLANIAESDAG   387

Qy  191 TLTCVATWKSLKARKSATVNLTV-------IRCPQDTGGGINIPGVLSSL           233
           : | :|       .| | | |      ||||||||||||||||||||
Db  388 VYTCHAA--NLAQQRRQDVNITVATVPSWLKKPQDSQLEEGKPGYLDCL           434
```

RESULT 11
A47299
ror-related receptor RTK - Pacific electric ray
N;Contains: protein-tyrosine kinase (EC 2.7.1.112)
C;Species: Torpedo californica (Pacific electric ray)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-May-2000
C;Accession: A47299
R;Jennings, C.G.; Dyer, S.M.; Burden, S.J.
Proc. Natl. Acad. Sci. U.S.A. 90, 2895-2899, 1993
A;Title: Muscle-specific trk-related receptor with a kringle domain defines a distinct o
A;Reference number: A47299; MUID:93219391; PMID:8385349
A;Accession: A47299
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-946 <JEN>
A;Cross-references: GB;L11311; NID:g2908057; PIDN:AAA49285.1; PID:g2908858
A;Experimental source: electric organ
A;Note: sequence extracted from NCBI backbone (NCBIN:128724, NCBIP:128726)
C;Superfamily: Torpedo ror-related receptor; immunoglobulin homology; kringle homology;
C;Keywords: ATP; glycoprotein; kringle; phosphotransferase; transmembrane protein; tyros
F;44-103/Domain: immunoglobulin homology <IM1>
F;137-195/Domain: immunoglobulin homology <IM2>
F;229-287/Domain: immunoglobulin homology <IM3>
F;464-542/Domain: kringle homology <KR3>
F;572-588/Domain: transmembrane #status predicted <TMN>
F;650-940/Domain: protein kinase homology <KIN>
F;658-666/Region: protein kinase ATP-binding motif
F;225,340,477,544/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match   5.5%;   Score 115.5; DB 1; Length 946;
Best Local Similarity 23.7%; Pred. No. 1.1;
Matches 62; Conservative 38; Mismatches 105; Indels 57; Gaps 12;

```
Qy    4 GAMENRDPPGSGSGNEVIEGPQNARVLKGSQOARFNC-TVSQGWKLIMWALSDMVVLSVRP  62
                          |  |:|| |  || |   |||.|  ::| |||||||||||||||||
Db  115 GALQVKMKP----KIIRPPTDVRALLGSKVVLPCSTMGNPKPAISWFKDETALKNDQP    168

Qy   63 MEPIIINDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNS---RLHGSAYLTVQV  119
          :  :                              |:.   | ||     |||| |||||
Db  169 RTSVL--------ESGN---LRIRNVQLEDAGYRCLARNSLGFEYSRSAALEVQV      213

Qy  120 MGELF-IPSVNLLVAENE---PCEVTCLPSHWTWLPDISW-ELGLLVSHSSYYFVPEPSD  174
           |                   |||||||||||||||||||  |:|||||||||||||||
```

Db 214 SARIVKAPTSQNYSYGSEVILQCKATGFP------IPTIKWLENGRAVPKGSIQNRIKGEV 268
Qy 175 LQSAVSILALTPQSNGTLLTCVATWK----SLKARKSATVNLTIVIRCPQDTGGGINIPGVL 230
Db 269 MESRLRVVTRP--SLFTCLITNKHNEGSTTAKATATALDIKEWRLYKG------ 314

Qy 231 SSLPSLGFSLPTWGKVGLGLIAG 252
Db 315 ----DLGYCSTYRGEVCQGLLG 332

RESULT 12
T18975
hypothetical protein C06A1.6 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T18975
R;McMurray, A.
submitted to the EMBL Data Library, June 1995
A;Reference number: Z19054
A;Accession: T18975
A;Molecule type: DNA
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Residues: 1-152 <WIL>
A;Cross-references: EMBL;Z49886; PIDN:CAA90055.1; GSPDB:GN00020; CESP:C06A1.6
A;Gene: CESP:C06A1.6
C;Genetics:
A;Experimental source: clone C06A1
A;Gene: CESP:C06A1.6
A;Map position: 2
A;Introns: 22/3

Query Match   5.5%;   Score 114.5; DB 2; Length 152;
Best Local Similarity 24.28%; Pred. No. 0.18;
Matches 29; Conservative 16; Mismatches 32; Indels 43; Gaps 5;

```
Qy  180 SILALTPQSNGTLLTCVATWKSLKARKSATVNLTVIR------CPQDTGGGINIPGVLSS   232
Db   16 TIVSISQVESGVLPVSTTEISIISSETSIIGSSRVKRNGGCGCGCCGGGGG--------    66

Qy  233 LPSLGFSLPTWGKVGLGLAGTMLLTFPTCTLTIRCCCCRRRCCGCN----CCCRCFCCR   287
Db   67 ----------GGCGCCCCRPKC-CTCCRTCCCTRCCTCCR                     104
```

RESULT 13
S20900
titin - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 18-Jun-1999
C;Accession: S20900
R;Labeit, S.; Gautel, M.; Lakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A;Title: Towards a molecular understanding of titin.
A;Reference number: S20897; MUID:92258380; PMID:1582406
A;Accession: S20900
A;Molecule type: mRNA
A;Status: nucleic acid sequence not shown; translation not shown
A;Residues: 1-531 <LAB>
A;Cross-references: EMBL;X64700; NID:g54807; PIDN:CAA45941.1; PID:g54808
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1992
C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology
C;Keywords: muscle

Query Match   5.5%;   Score 114.5; DB 2; Length 531;
Best Local Similarity 23.4%; Pred. No. 0.67;
Matches 62; Conservative 36; Mismatches 98; Indels 69; Gaps 11;

```
Qy    3 AGAMENRDPPGSGSGNEVIEGPQNARVLKGSQOARFNCTVSQGWKLIMWALSDMVVLSVRP  62
Db   93 AGVGEHADVPGPVMVEEKLEAPD---IDLDLELRKVINIRAGGSLRLF-----VPIKGRP  144

Qy   63 MEPI----ITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAYLTVQ 118
```

```
Db   274 LRGEEVI------------------QLRSKKYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNEN 318
Qy   109 LHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTWLPDISW-ELGLLVSHSSYY 167
Db   319 ISASAELTVLVPPWFLNHPSNLYAYESMDIEFECTVSGKP-VPTVNWMKNGDVVIPSDYF 377
Qy   168 FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIP 227
Db   378 QIVGGSNLR----ILGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS------ 426
Qy   228 GVLSSLP 234
Db   427 -VLPSAP 432
```

RESULT 2
PCT-US94-05277-2
Sequence 2, Application PC/TUS9405277

GENERAL INFORMATION:
    APPLICANT: Bruskin, Arthur
    APPLICANT: Jarosz, David E.
    APPLICANT: Johnson, Karen
    APPLICANT: Kinzler, Kenneth W.
    APPLICANT: Zabrecky, James R.
    APPLICANT: Vogelstein, Bert
    TITLE OF INVENTION: Antibodies Specific for DCC Gene Product
    NUMBER OF SEQUENCES: 2
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Banner, Birch, McKie & Beckett
        STREET: 1001 G Street, N.W.
        CITY: Washington
        STATE: D.C.
        COUNTRY: USA
        ZIP: 20001
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: PCT/US94/05277
        FILING DATE:
        CLASSIFICATION:
    ATTORNEY/AGENT INFORMATION:
        NAME: Kagan, Sarah A.
        REGISTRATION NUMBER: 32,141
        REFERENCE/DOCKET NUMBER: 01107.42709
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 202.508.9100
        TELEFAX: 202.508.9299
        TELEX: 197430 BBMB UT
    INFORMATION FOR SEQ ID NO: 2:
        SEQUENCE CHARACTERISTICS:
        LENGTH: 1447 amino acids
        TYPE: amino acid
        TOPOLOGY: linear
        MOLECULE TYPE: protein

Query Match   6.5%;   Score 136;   DB 5;   Length 1447;
Best Local Similarity   25.1%;   Pred. No. 0.0018;
Matches   62;   Conservative   34;   Mismatches 101;   Indels 50;   Gaps 10;

```
Qy     9 RDPGGSGSGNEV-----------IEGPQNARVLKGSQARFNCTVSQGW--KLIMW 50
Db   215 RNPASSRTGNEAEVRILSDPGLHRQLYFLQRPSNVVAIEGKDAVLECCVS-GYPPPSFTW 273
Qy    51 ALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC--SLQNSR 108
Db   274 LRGEEVI-------------QLRSKKYSLLGG--SNLLISNVTDDDSGMYTCVVTYKNEN 318
Qy   109 LHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTWLPDISW-ELGLLVSHSSYY 167
```

US-08-374-834-16

```
Db   319 ISASAELTVLVPPWFLNHPSNLYAYESMDIEFECTVSGKP-VPTVNWMKNGDVVIPSDYF 377
Qy   168 FVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLKARKSATVNLTVIRCPQDTGGGINIP 227
Db   378 QIVGGSNLR----ILGVVKSDEGFYQCVAENEAGNAQTSAQLIVPKPAIPSSS------ 426
Qy   228 GVLSSLP 234
Db   427 -VLPSAP 432
```

RESULT 3
US-08-374-834-16
Sequence 16, Application US/08374834
Patent No. 5656473

GENERAL INFORMATION:
    APPLICANT: Valenzuela, et al.
    TITLE OF INVENTION: NOVEL TYROSINE KINASE RECEPTOR
    NUMBER OF SEQUENCES: 17
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Regeneron Pharmaceuticals, Inc.
        STREET: 777 Old Saw Mill River Road
        CITY: Tarrytown
        STATE: New York
        COUNTRY: USA
        ZIP: 10591
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/374,834
        FILING DATE: 19-JAN-1995
        CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 08/095,658
        FILING DATE: 21-JUL-1993
    ATTORNEY/AGENT INFORMATION:
        NAME: Cobert, Robert J.
        REGISTRATION NUMBER: 36,108
        REFERENCE/DOCKET NUMBER: REG 190A
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: (914) 345-7400
        TELEFAX: (914) 345-7721
    INFORMATION FOR SEQ ID NO: 16:
        SEQUENCE CHARACTERISTICS:
        LENGTH: 869 amino acids
        TYPE: amino acid
        STRANDEDNESS:
        TOPOLOGY: unknown
        MOLECULE TYPE: protein

Query Match   6.3%;   Score 132.5;   DB 1;   Length 869;
Best Local Similarity   20.8%;   Pred. No. 0.0018;
Matches   55;   Conservative   89;   Mismatches 178;   Indels 105;   Gaps 20;

```
Qy     4 GAMENRDPPGSGSGSGNEVIEGPQNARVLKGSQARFNCTVSQGWK-LIMWALSDMVVLSVRP 62
Db   113 GALQVKMKP--------KITRPPINVKIIEGLKAVLPCTTMGNPKPSVSWIKGD-------- 158
Qy    63 MEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQNSRLHGSAY-LTVQVMG 121
Db   159 -SPLRENSRIAVLE------SGSLRIHNVQKEDAGQYRCVAKNSL--GTAYSKVVKLEV 208
Qy   122 ELFIPSVNLVVAENEP------CEVTCLPSHWTWLPDISW-ELGLLVSHSSYYFVPEP 172
Db   209 EVFARILRAPESHNVTFGSFVILHCTATGIP----VPTITWIENGNAVSSGCSIQESVKD 263
Qy   173 SDLQSAVSILALTPQSNGTLTCVAT-----WKSLKARKSATVNLTVIRCPQDTGGG 223
```

OM protein - protein search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 18:14:43 ; Search time 14.0485 Seconds
(without alignments)
808.430 Million cell updates/sec

Title: US-09-729-264-4
Perfect score: 2088
Sequence: 1 MVAGAMENRDPGSGSGNEV..............HPQASFNLASPEKVSNTTVV 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	136	6.5	1447	4	US-09-041-886-25	Sequence 25, Appli
2	136	6.5	1447	5	PCT-US94-05277-2	Sequence 2, Appli
3	132.5	6.3	869	1	US-08-374-834-16	Sequence 16, Appl
4	132.5	6.3	869	2	US-08-644-271-29	Sequence 29, Appl
5	132.5	6.3	869	4	US-09-077-955-33	Sequence 33, Appl
6	128.5	6.2	1345	2	US-08-777-767-3	Sequence 3, Appl
7	123.5	5.9	340	4	US-09-651-200-2	Sequence 2, Appl
8	123.5	5.9	441	4	US-09-651-200-4	Sequence 4, Appl
9	122.5	5.9	534	4	US-09-651-200-6	Sequence 6, Appl
10	122.5	5.9	534	4	US-09-651-200-24	Sequence 24, Appl
11	122.5	5.8	478	5	US-08-633-148-2	Sequence 2, Appl
12	121.5	5.8	332	4	US-09-062-365-11	Sequence 11, Appl
13	120	5.7	1395	4	US-09-540-245A-15	Sequence 15, Appl
14	118	5.7	868	1	US-08-374-834-1	Sequence 1, Appl
15	118	5.7	868	2	US-08-644-271-1	Sequence 1, Appl
16	118	5.6	868	4	US-09-077-955-1	Sequence 1, Appl
17	116.5	5.6	318	2	US-08-633-148-4	Sequence 4, Appl
18	116.5	5.6	340	2	US-08-633-148-2	Sequence 2, Appl
19	115.5	5.6	860	5	PCT-US95-08493-19	Sequence 19, Appl
20	115.5	5.5	868	5	PCT-US95-08493-21	Sequence 21, Appl
21	115.5	5.5	946	5	PCT-US95-08493-13	Sequence 13, Appl
22	113	5.4	325	4	US-09-651-200-20	Sequence 20, Appl
23	112	5.4	689	4	US-09-499-964-1	Sequence 1, Appl
24	109	5.2	421	2	US-08-659-984A-1	Sequence 1, Appl
25	109	5.2	421	2	US-08-660-531-1	Sequence 1, Appl
26	109	5.2	444	2	US-08-659-984A-5	Sequence 5, Appl
27	109	5.2	444	4	US-08-660-531-5	Sequence 5, Appl
28	107.5	5.1	362	1	US-08-415-751-6	Sequence 6, Appli
29	107	5.1	1297	4	US-09-540-245A-17	Sequence 17, Appli
30	106	5.1	801	1	US-07-906-349A-6	Sequence 6, Appli
31	105	5.1	285	3	US-08-482-085B-20	Sequence 20, Appli
32	104	5.0	319	1	US-08-597-495B-22	Sequence 22, Appl
33	104	5.0	319	4	US-09-068-051A-22	Sequence 22, Appli
34	104	5.0	319	4	US-09-336-536-67	Sequence 67, Appli
35	104	5.0	319	4	US-09-254-465A-6	Sequence 6, Appli
36	104	5.0	690	4	US-08-935-433-2	Sequence 2, Appli
37	104	5.0	365	4	US-08-928-383B-26	Sequence 26, Appli
38	103.5	5.0	879	1	US-08-554-612C-1	Sequence 1, Appli
39	102.5	5.0	890	1	US-08-445-640-2	Sequence 2, Appli
40	102.5	4.9	890	1	US-09-553-132-2	Sequence 2, Appli
41	102.5	4.9	890	3	US-08-170-558-2	Sequence 2, Appli
42	102.5	4.9	890	3	US-08-447-314-2	Sequence 2, Appli
43	102.5	4.9	890	3	US-08-445-461-2	Sequence 2, Appli
44	102.5	4.9	911	1	US-08-286-305A-1	Sequence 1, Appli
45	102.5	4.9	911	2	US-08-441-104A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-041-886-25
Sequence 25, Application US/09041886
Patent No. 6235872

GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: Rabizadeh, Sharroz
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-9049
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1447 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-041-886-25

Query Match   6.5%;
Best Local Similarity   25.18%;
Matches 62 Conservative 34;

Score 136; DB 4; Length 1447;
Pred. No. 0.0018;
Mismatches 101; Indels 50; Gaps 10;

```
Qy      9 RDPPGSGSGNEV-----------------IEGPQNARVLKGSQARFNCTVSQGW--KLIMW 50
Db    215 RNPASSRTGNEAEVRILSDPGLIHRQLYFLQRPSNVVAIEGKDAVLECCVS-GYPPPSFTW 273

Qy     51 ALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNYEPSDSGNIRC--SLQNSR 108
```

```
Qy    873 GCTGCGTGCGCTGCGGCCCGGCCGCGTCGTTGTTGTTGTTGGGCTGCGAACTGCGTGCGTGTTGTTGTTTCT 932
          |||||||||| | || | | ||||| |||||||  ||| |||||||||| ||| |||| | |||| | ||
Db     67 GCTGCGCTGCGCCCGCGCCTGCTGCTGCTGCTGCGCTGTTGCTGCGTGTTGCATCTGCAGCTGTTTGCTGGTGTTTGCTGTTGCTGTTGCT 8
```

```
Qy    933 GCTG  936
          | | |
Db      7 GGTG  4
```

RESULT 13
US-08-145-617-5/c
Sequence 5, Application US/08145617
Patent No. 5766847

GENERAL INFORMATION:
;   APPLICANT: Jackle, Herbert
;   APPLICANT: Tautz, Diethard
;   TITLE OF INVENTION: PROCESS FOR ANALYZING LENGTH
;   TITLE OF INVENTION: POLYMORPHISMS IN DNA REGIONS
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
;   STREET: 301 N. Washington Street, P.O. Box 747
;   CITY: Falls Church
;   STATE: Virginia
;   COUNTRY: United States of America
;   ZIP: 22046
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/145,617
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/681,494
;   FILING DATE: 10-JUN-1991
;   APPLICATION NUMBER: DE P3834636.2
;   FILING DATE: 11-OCT-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Svensson, Leonard R.
;   REGISTRATION NUMBER: 30,330
;   REFERENCE/DOCKET NUMBER: 147-122PCT
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 703-241-1300
;   TELEFAX: 703-241-2848
;   TELEX: 248345
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 379 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)

Query Match 67.4%; Score 41.2; DB 1; Length 379;
Best Local Similarity 67.4%; Pred. No. 0.027;
Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```
Qy    851 CCGACGTGTACATCTTACAATACGGCTCGTGCGCCGGCCGCCGTTCGTTGTTGTTGTGGGCTGCAAC 910
          ||| |    |||||| | ||||| |||||| || ||  |||||||||||| |||||  ||
Db    214 CCGAACTCCAGGCGCCTCCGAGTTGCTGCTGCTTGCTGCTGTTGCTGCTGTTTGCTGTTGCTGCT 155
```

```
Qy    911 TGCTGCGTGCCGTTGTTGTTGTTTCGCCTG  936
          |||||||| |||||||| ||||| | ||
Db    154 TGCTGCTGCTGATTGTTGCTGCTGTTG 129
```

RESULT 14
US-08-185-432-3/c
Sequence 3, Application US/08185432
Patent No. 5750652

GENERAL INFORMATION:
;   APPLICANT: Artavanis-Tsakonas, Spyridon
;   APPLICANT: Busseau, Isabelle
;   APPLICANT: Diederich, Robert J.
;   APPLICANT: Xu, Tian
;   APPLICANT: Matsuno, Kenji
;   TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
;   TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PENNIE & EDMONDS
;   STREET: 1155 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10036-2711
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/185,432
;   FILING DATE: 21-JAN-1994
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Misrock, S. Leslie
;   REGISTRATION NUMBER: 18,872
;   REFERENCE/DOCKET NUMBER: 7326-006
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 790-9090
;   TELEFAX: (212) 869-8864/9741
;   TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3771 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: unknown
;   MOLECULE TYPE: cDNA
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 345..2558
US-08-185-432-3

Query Match 66.3%; Score 41; DB 1; Length 3771;
Best Local Similarity 66.3%; Pred. No. 0.1;
Matches 59; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

```
Qy    873 GCTGCTGCGTGCGCTGCGGCCCGGCCGCGTCGTTGTTGTGGCTGCAACTGCGTGCGTGCCGTTGTTGTTGTTTCT 932
          ||||||||||| | | | ||||| |||||| | ||||| |||||| ||| |||| ||||||||  |||||| ||
Db   1875 GCTGCGTGCTGCGCCCTGTTGCTGCGTGTTGCATCTGCAGCTGTTTGCTGCTGTTGCTGTTGCT 1816
```

```
Qy    933 GCTGTAGAAGAAAAAGAGGAGATTTCGTATT 961
          |||||| |   |  | | |||| |||
Db   1815 GCTGTTGCTGCTGCGCGGATGCAGTGGTAAT 1787
```

RESULT 15
US-08-185-432-23
Sequence 23, Application US/08185432
Patent No. 5750652

GENERAL INFORMATION:
;   APPLICANT: Artavanis-Tsakonas, Spyridon
;   APPLICANT: Busseau, Isabelle
;   APPLICANT: Diederich, Robert J.
;   APPLICANT: Xu, Tian
;   APPLICANT: Matsuno, Kenji

PCT-US92-11107-12

```
Query Match              3.4%;    Score 42.6;  DB 5;   Length 3292;
Best Local Similarity    64.9%;   Pred. No. 0.033;
Matches 63;   Conservative 0;   Mismatches 34;   Indels 0;   Gaps 0;

Qy    869  ATACGCTGCTGCTCGCCGCCGTCGTTGTTGTGGCGTCGCAACTCCTGCCGTTGTTCT   928
           ||||||| |||| ||||||||||||||||||| ||||||||||||||||||||||||
Db    818  AGATGATGATGTTGTTGTTGGTTGTTGAAGTTGTGTTGTTGCAACTGGTTGTTGCGTGTTGTAGC  759

Qy    929  TTCTGCTGTGTAGAAGAAAAAGAGGATTTCGTATTCAAT  965
           |||||||| |||||||||||||||||||||| |||||||
Db    758  TGCTGCTGTTGCTGCAACTGCCTGTTGTTGTTGTAATT  722
```

RESULT 9
US-780-049-18/c
; Sequence 18, Application US/09780049
; Patent No. 6465250
; GENERAL INFORMATION:
;   APPLICANT: Brett P. Monia
;   APPLICANT: Jacqueline Wyatt
;   TITLE OF INVENTION: ANTISENSE MODULATION OF PROTEIN PHOSPHATASE 2 CATALYTIC SUBUNIT
;   FILE REFERENCE: RTS-0134
;   CURRENT APPLICATION NUMBER: US/09/780,049
;   CURRENT FILING DATE: 2001-02-09
;   NUMBER OF SEQ ID NOS: 96
;   SEQ ID NO 18
;   LENGTH: 40000
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
US-09-780-049-18

RESULT 10
US-09-817-180-1
; Sequence 1, Application US/09817180
; Patent No. 6340584
; GENERAL INFORMATION:
;   APPLICANT: GAN, Weinlu et al.
;   TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;   TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;   TITLE OF INVENTION: THEREOF
;   FILE REFERENCE: CL001183
;   CURRENT APPLICATION NUMBER: US/09/817,180
;   CURRENT FILING DATE: 2001-03-27
;   NUMBER OF SEQ ID NOS: 4
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 1
;   LENGTH: 2674
;   TYPE: DNA
;   ORGANISM: Human
US-09-817-180-1

```
Query Match              3.4%;    Score 42;    DB 4;   Length 2674;
Best Local Similarity    51.6%;   Pred. No. 0.044;
Matches 96;   Conservative 0;   Mismatches 90;   Indels 0;   Gaps 0;

Qy    882  GCTGCCGCGCCGTCGTTGTTGTTGGGCTGCAACTGCTGCTCGCCGTTGTTGTTTCTGCTGTAGAA  941
```

```
Query Match              3.4%;    Score 42.6;  DB 4;   Length 40000;
Best Local Similarity    67.4%;   Pred. No. 0.13;
Matches 60;   Conservative 0;   Mismatches 29;   Indels 0;   Gaps 0;

Qy    28     TCAGTTTCCTTAGGCTGCCATAACAAAAGCACCATAACCTGGTGGCTTAGANACAATTGGAAAG   87
             ||||| |||  ||||||||| |||||||||||||||||  |||||||||||||||||  |||||
Db    24176  TCTTTTGCTPAAGGCTGCCACACAACAAAGCATCAGAGAAIAGGTGGCTTAAACAACACAGAAAT  24117
```

```
Qy    88     GCAATTGCTCACGGTTCCACAGAAGCTGTAG  116
             || ||||||||||| |||||||||| |||| |
Db    24116  TTACTTCCTCCACTGTTCCAGAGGCAGAAG  24088
```

RESULT 11
US-09-043-303-12/c
; Sequence 12, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
;   APPLICANT: TSUJI, Shoji
;   APPLICANT: SANPEI, Kazujiro
;   TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
;   TITLE OF INVENTION: Primers Therefor
;   FILE REFERENCE: 0760-0241P
;   CURRENT APPLICATION NUMBER: US/09/043,303
;   CURRENT FILING DATE: 1998-05-18
;   EARLIER APPLICATION NUMBER: PCT/JP96/01999
;   EARLIER FILING DATE: 1996-07-18
;   NUMBER OF SEQ ID NOS: 17
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 12
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-043-303-12

```
Qy    1062  CAGAAA  1067
            | |||
Db    2640  AAAAAA  2645
```

```
Db    2460  GATGCAGCGCGGCGTGCGTGTCCTCTCTCTGTTGTCCTGCTGCTGCTGCCAGGGCCTTCCTCTTCCGGGCAG  2519
            |||| ||    || |||||||||  ||||||||| | || | ||||| | || | |||| |
Qy    942   GAAAAGAGGAGATTTCGTATTCAAPTTTCAAAAGAAAATCTGAAAAAAGAGAGAGACAAACAAAG  1001
            |  |||| ||||  ||| ||| |||  |  ||  ||  || ||  |||   |||  |||| |||
Db    2520  AAACAATAAAACCACTTGTCCCACTGAAAAAAaaaaaaaaaaaaaaaaaaaaaaaa  2579

Qy    1002  AAACTGAGACAGAAAGTGGAAATGAAAACTCCGGCTACAATTCAGATGAACAAAAGACCA  1061
            || |||    ||  ||||| || ||   |||  |||||| |||| |||  | ||||
Db    2580  AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2639
```

RESULT 12
US-09-043-303-7/c
; Sequence 7, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
;   APPLICANT: TSUJI, Shoji
;   APPLICANT: SANPEI, Kazujiro
;   TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
;   TITLE OF INVENTION: Primers Therefor
;   FILE REFERENCE: 0760-0241P
;   CURRENT APPLICATION NUMBER: US/09/043,303
;   CURRENT FILING DATE: 1998-05-18
;   EARLIER APPLICATION NUMBER: PCT/JP96/01999
;   EARLIER FILING DATE: 1996-07-18
;   NUMBER OF SEQ ID NOS: 17
;   SOFTWARE: PatentIn Ver. 2.0
;   SEQ ID NO 7
;   LENGTH: 203
;   TYPE: DNA
;   ORGANISM: p-2093 plasmid
US-09-043-303-7

```
Query Match              3.4%;    Score 41.6;  DB 4;   Length 78;
Best Local Similarity    78.1%;   Pred. No. 0.0088;
Matches 50;   Conservative 0;   Mismatches 14;   Indels 0;   Gaps 0;

Qy    873  GCTGCTGCTGCTGCGCGCCGGTCGTTGTTGTTGGGCTGCAACTGCTGCTCGCCGTTGTTGTTTCT  932
           ||||||||| |||| |||||| ||||||||||||||  || ||||||||||||||||| ||||||
Db    67   GCTGCTGCTGCTGCTTGGCTTGCTGTCTGCTGCTGCTGCTGCTGCTGCTCGCCGTTGTTGTTCGCTGCT  8
```

```
Qy    933  GCTG  936
           ||||
Db    7    GCTG  4
```

MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
IMMEDIATE SOURCE:
CLONE: lambda yPt
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1626
US-08-258-442-12

Qy 869 ATACGCTGCTGCTGCTGCTGCCGCCGCCGTGCTGCTGTTTGTTGGCCTGCAACTGCTTGCCTCCGGTTGTTGT 928
       | ||||||||| | || || | | || |||||| |||||| || | | ||||||||| ||| || | ||||
Db 818 AGATGATGATGTGTTGTTGTTGAAGTTGTTGTTGTTGGTTGCAACTGGTGTTGTTGCTTGGTTGTTGTAGC 759

Qy 929 TTCTGCTGTAGAAGAAAAAGAGGATTTCGTATTCAAT 965
       | ||||||| ||| | ||| |||| |||||| |
Db 758 TGCTGCTGTTGCTGCGACAACTGCTGTTGTTGTTGTAATT 722

Query Match 3.4%; Score 42.6; DB 1; Length 3292;
Best Local Similarity 64.9%; Pred. No. 0.033;
Matches 63; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

RESULT 7
US-08-328-809-7/c
Sequence 7, Application US/08328809
Patent No. 5705334
GENERAL INFORMATION:
APPLICANT: Lippard, Stephen J.
APPLICANT: Essigmann, John M.
APPLICANT: Donahue, Brian A.
APPLICANT: Toney, Jeffrey H.
APPLICANT: Bruhn, Suzanne L.
APPLICANT: Pil, Pieter M.
APPLICANT: Brown, Steven
APPLICANT: Kellett, Patti
TITLE OF INVENTION: Uses For DNA Structure-Specific
TITLE OF INVENTION: Recognition Proteins
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patent Administrator, Testa, Hurwitz & Thibeault
STREET: 53 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,809
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Fenton, Gillian M.
REGISTRATION NUMBER: 36,508
REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-248-7000
TELEFAX: 617-248-7100
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 3292 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
IMMEDIATE SOURCE:
CLONE: lambda yPt

FEATURE:
NAME/KEY: CDS
LOCATION: 1..1626
US-08-328-809-7

Qy 869 ATACGCTGCTGCTGCTGCTGCCGCCGCCGTGCTGCTGTTTGTTGGCTGCAACTGCTTGCCTCCGGTTGTTGT 928
       | ||||||||| | || || | | || |||||| |||||| || | | ||||||||| ||| || | ||||
Db 818 AGATGATGATGTGTTGTTGTTGAAGTTGTTGTTGTTGGTTGCAACTGGTGTTGTTGCTTGGTTGTTGTAGC 759

Qy 929 TTCTGCTGTAGAAGAAAAAGAGGATTTCGTATTCAAT 965
       | ||||||| ||| | ||| |||| |||||| |
Db 758 TGCTGCTGTTGCTGCGACAACTGCTGTTGTTGTTGTAATT 722

Query Match 3.4%; Score 42.6; DB 1; Length 3292;
Best Local Similarity 64.9%; Pred. No. 0.033;
Matches 63; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

RESULT 8
PCT-US92-11107-12/c
Sequence 12, Application PCT/US9211107
GENERAL INFORMATION:
APPLICANT: Donahue, Brian A.
APPLICANT: Toney, Jeffrey H.
APPLICANT: Bruhn, Suzanne L.
APPLICANT: Pil, Pieter M.
APPLICANT: Brown, Steven
APPLICANT: Kellett, Patti
APPLICANT: Lippard, Stephen J.
APPLICANT: Essigmann, John M.
TITLE OF INVENTION: DNA Structure Specific Recognition
TITLE OF INVENTION: Protein and Uses Therefor
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: 2 Militia Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/11107
FILING DATE: 19921218
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/539,906
FILING DATE: 18-JUN-1990
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4787AAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3292 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Saccharomyces cerevisiae
IMMEDIATE SOURCE:
CLONE: lambda yPt
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1626

Left column (RESULT 5)

```
;
;     TELEPHONE: (619) 535-9001
;     TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 12:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 1776 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
;     FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 36..1116
US-09-041-886-12

Query Match     3.5%;  Score 42.8;  DB 4;  Length 1776;
Best Local Similarity 61.8%;  Pred. No. 0.024;
Matches  68;  Conservative  0;  Mismatches  42;  Indels  0;  Gaps  0;

Qy   873 GCTGCTGTGCTGCCGCCGTCGTGTTGTTGTGCGTCGCCGTTGTTGTTTCT 932
         ||||||||||||||||| |||||||||||| | ||| |||| || ||||
Db   975 GCTGCTGCGCTGCCTGCGTCGCTGCTGCCTGCTGCTGCGTCTGCTGTTCT 916

Qy   933 GCTGTAGAAGAAAAAAGAGGATTTCGTATTCAATTTCAAAAGAAATCTGAA 982
         ||||  | ||||| | |||||||  |  |||   ||||||| |||  |||
Db   915 GCTGCTGTTTTTCAAAGTAGGCTTCTCGTCTCTTCCGAAGCTCTTCTGAA 866
```

```
RESULT 5
US-07-814-964-12/c
Sequence 12, Application US/07814964
Patent No. 5359047
GENERAL INFORMATION:
    APPLICANT: Donahue, Brian A.
    APPLICANT: Toney, Jeffrey H.
    APPLICANT: Bruhn, Suzanne L.
    APPLICANT: Pil, Pieter M.
    APPLICANT: Brown, Steven
    APPLICANT: Kellett, Patti
    APPLICANT: Essigmann, John M.
    APPLICANT: Lippard, Stephen J.
    TITLE OF INVENTION: DNA Structure Specific Recognition
    TITLE OF INVENTION: Protein and Uses Therefor
    NUMBER OF SEQUENCES: 13
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
        STREET: 2 Militia Drive
        CITY: Lexington
        STATE: MA
        COUNTRY: USA
        ZIP: 02173
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/07/814,964
        FILING DATE: 19911226
        CLASSIFICATION: 435
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/539,906
        FILING DATE: 18-JUN-1990
    ATTORNEY/AGENT INFORMATION:
        NAME: Granahan, Patricia
        REGISTRATION NUMBER: 32,227
        REFERENCE/DOCKET NUMBER: MIT-4787AAA
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 617-861-6240
        TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 3292 base pairs
        TYPE: NUCLEIC ACID
```

Right column (RESULT 6)

```
;
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;     MOLECULE TYPE: DNA (genomic)
;     ORIGINAL SOURCE:
;     ORGANISM: Saccharomyces cerevisiae
;     IMMEDIATE SOURCE:
;     CLONE: lambda ypt
;     FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..1626
US-07-814-964-12

Query Match     3.4%;  Score 42.6;  DB 1;  Length 3292;
Best Local Similarity 64.9%;  Pred. No. 0.033;
Matches  63;  Conservative  0;  Mismatches  34;  Indels  0;  Gaps  0;

Qy   869 ATACCGCTGCGCTGCCTGCCCGCGTCGTTGTTGTTGTTGGCTGCAACTGC 928
         | |||  | ||||||| ||  | ||||||||||||||| || ||| |||
Db   818 AGATGATGATGTGTTGTTGAAGTGTTGTTGTTGTTGCAACTGGTGTTGTGT 759

Qy   929 TTCTGCTGTAGAAGAAAAAGAGGATTTCGTATTCAAT 965
         |||| || ||  | |||| ||||||||| ||  | |
Db   758 TGCTGCTGTTGCTGCGAACTGCTGTTGTTGTTGTAATT 722
```

```
RESULT 6
US-08-258-442-12/c
Sequence 12, Application US/08258442
Patent No. 5670621
GENERAL INFORMATION:
    APPLICANT: Donahue, Brian A.
    APPLICANT: Toney, Jeffrey H.
    APPLICANT: Bruhn, Suzanne L.
    APPLICANT: Pil, Pieter M.
    APPLICANT: Brown, Steven
    APPLICANT: Kellett, Patti
    APPLICANT: Essigmann, John M.
    APPLICANT: Lippard, Stephen J.
    TITLE OF INVENTION: DNA Structure Specific Recognition
    TITLE OF INVENTION: Protein and Uses Therefor
    NUMBER OF SEQUENCES: 13
    CORRESPONDENCE ADDRESS:
        ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
        STREET: 2 Militia Drive
        CITY: Lexington
        STATE: MA
        COUNTRY: USA
        ZIP: 02173
    COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk
        COMPUTER: IBM PC compatible
        OPERATING SYSTEM: PC-DOS/MS-DOS
        SOFTWARE: Patentin Release #1.0, Version #1.25
    CURRENT APPLICATION DATA:
        APPLICATION NUMBER: US/08/258,442
        FILING DATE:
        CLASSIFICATION: 530
    PRIOR APPLICATION DATA:
        APPLICATION NUMBER: US 07/539,906
        FILING DATE: 18-JUN-1990
    ATTORNEY/AGENT INFORMATION:
        NAME: Granahan, Patricia
        REGISTRATION NUMBER: 32,227
        REFERENCE/DOCKET NUMBER: MIT-4787AAA
    TELECOMMUNICATION INFORMATION:
        TELEPHONE: 617-861-6240
        TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
    SEQUENCE CHARACTERISTICS:
        LENGTH: 3292 base pairs
        TYPE: nucleic acid
        STRANDEDNESS: double
        TOPOLOGY: linear
```

STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/531,927B
FILING DATE: 21-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP H6-251600
FILING DATE: 21-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: ATH95-01A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 325 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: modified_base
LOCATION: 145
FEATURE:
NAME/KEY: modified_base
LOCATION: 194
US-08-531-927B-3

Query Match   3.9%;   Score 48.2;   DB 2;   Length 325;
Best Local Similarity   57.8%;   Pred. No. 0.00024;
Matches 100; Conservative 2; Mismatches 70; Indels 1; Gaps 1;

```
Qy  857 TGTACTCTTACAATACGCTGCTGCCCGCGTCGTTGTTGGCTGCAACTGCTGC 916
            ||| || |||| ||||||||||||| ||||||||||||| | |||||
Db  210 TGTCCTGATTAGGTCCCSCTGCGTCGTCGTGCTGTCGTGCTGCGGGGTGCTG 151

Qy  917 TGCCGTTGTGTTGTTTCTCGCTGTGTAGAA-GAAAAGAGGA 975
            ||| ||| || ||  |||||  ||||  |||| |||
Db  150 TGCTGYTCCTGCTTTTGCTGCGTCTCGAAACATTCAAAAGTGA 91

Qy  976 ATCTGAAAAAGAGAACAAAAGAAAAAGAAAACTGAGACAGAAAGTGGAAATGAAA 1028
            |  ||| | | ||  |||  ||   | || |||   |  |||   | || |
Db   90 AAACTTAAAGATAATATACACCATGAGAAAAACTATTCATAAGGAAAAATACA 38
```

RESULT 3
US-08-531-927B-1/c
Sequence 1, Application US/08531927B
Patent No. 5840491
GENERAL INFORMATION:
APPLICANT: Kakizuka, Akira
TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
Patent No. 5840491
TITLE OF INVENTION: Disease Gene and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/531,927B
FILING DATE: 21-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP H6-251600
FILING DATE: 21-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: ATH95-01A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1776 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 36..1115
US-08-531-927B-1

Query Match   3.58;   Score 42.8;   DB 2;   Length 1776;
Best Local Similarity   61.8%;   Pred. No. 0.021;
Matches 68; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Qy  873 GCTGCTGCTGCTGCGCCGCGGCGTGCGGTTGTTGTTGGCTGCAACTGCTGCTGCCGGGTTGTTGTTTCT 932
            ||||| ||||| | || || ||| ||  |||| | ||||| |||||| |||||| ||| ||||| |
Db  975 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCGTCGTGCTGCTGCTGCGTCGTGCTGCTGTTGCTGCTTTT 916

Qy  933 GCTGTAGAAGAAAAAGAGGGATTTCGTATTCAATTTCAAAAGAAATCTGAA 982
            ||||   || ||    |  ||  | | ||    |||   |  |||| |||||
Db  915 GCTGCTGTTTTCAAAGTAGGCTTCTCGTTCTTCCGAAGCTCTTCTGAA 866
```

RESULT 4
US-09-041-886-12/c
Sequence 12, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: Rabizadeh, Sharroz
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2003, 23:25:03 ; Search time 51.9118 seconds
                                    (without alignments)
                                    7325.486 Million cell updates/sec

Title: US-09-729-264-5
Perfect score: 1240
Sequence: 1 aggtgtgagtccagccaaca..........gtaatacaactgtagtatag 1240

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA.*
            1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
            2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
            3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
            4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
            5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Match Length	DB ID	Description
C	1	51.8	4.2	397	US-09-253-691-3	Sequence 3, Appli
C	2	48.2	3.9	325	US-08-531-927B-3	Sequence 3, Appli
C	3	42.8	3.5	1776	US-08-531-927B-1	Sequence 1, Appli
C	4	42.8	3.4	1776	US-09-041-886-12	Sequence 12, Appli
C	5	42.6	3.4	3292	US-07-814-964-12	Sequence 12, Appli
C	6	42.6	3.4	3292	US-08-258-442-12	Sequence 12, Appli
C	7	42.6	3.4	3292	US-08-328-809-7	Sequence 7, Appli
C	8	42.6	3.4	3292	PCT-US92-11107-12	Sequence 12, Appli
C	9	42.6	3.4	40000	US-09-780-049-18	Sequence 18, Appli
C	10	42	3.4	2674	US-09-817-180-1	Sequence 1, Appli
C	11	41.6	3.4	78	US-09-043-303-12	Sequence 12, Appli
C	12	41.6	3.4	203	US-09-043-303-7	Sequence 7, Appli
C	13	41.2	3.3	379	US-08-145-617-5	Sequence 5, Appli
C	14	41	3.3	3771	US-08-185-432-23	Sequence 23, Appli
C	15	41	3.3	3771	US-08-185-432-3	Sequence 3, Appli
C	16	41	3.3	5063	US-08-185-432-1	Sequence 1, Appli
C	17	40.8	3.3	6177	US-08-479-913E-1	Sequence 1, Appli
C	18	40.8	3.3	7257	US-09-091-042A-1	Sequence 1, Appli
C	19	40.4	3.3	484	Patent No. 5496550	Patent No. 5496550
C	20	40.4	3.3	10348	US-08-457-273B-41	Sequence 41, Appl
C	21	40.4	3.3	10348	US-08-556-419-13	Sequence 13, Appl
C	22	40.4	3.3	10348	US-09-041-886-14	Sequence 14, Appli
C	23	40.4	3.3	10366	US-08-453-265-5	Sequence 5, Appli
C	24	40.4	3.3	10366	US-08-246-982A-5	Sequence 5, Appli
C	25	40	3.2	75	US-09-043-303-10	Sequence 10, Appl
C	26	40	3.2	78	US-09-043-303-11	Sequence 11, Appl
C	27	40	3.2	154	US-08-469-802B-6	Sequence 6, Appli
C	28	40	3.2	154	US-08-267-803B-6	Sequence 6, Appli
C	29	40	3.2	165	US-09-043-303-17	Sequence 17, Appl
C	30	40	3.2	168	US-08-469-802B-4	Sequence 4, Appli
C	31	40	3.2	168	US-08-267-803B-4	Sequence 4, Appli
C	32	40	3.2	171	US-08-469-802B-5	Sequence 5, Appli
C	33	40	3.2	171	US-08-267-803B-5	Sequence 5, Appli
C	34	40	3.2	195	US-08-469-802B-2	Sequence 2, Appli
C	35	40	3.2	195	US-08-267-803B-2	Sequence 2, Appli
C	36	40	3.2	234	US-08-469-802B-3	Sequence 3, Appli
C	37	40	3.2	234	US-08-267-803B-3	Sequence 3, Appli
C	38	40	3.2	270	US-09-146-054-8	Sequence 8, Appli
C	39	40	3.2	270	US-09-664-977A-8	Sequence 8, Appli
C	40	40	3.2	477	US-09-135-994-1	Sequence 1, Appli
C	41	40	3.2	506	US-08-469-802B-7	Sequence 7, Appli
C	42	40	3.2	506	US-08-267-803B-7	Sequence 7, Appli
C	43	40	3.2	623	US-09-043-303-5	Sequence 5, Appli
C	44	40	3.2	3366	US-08-469-802B-1	Sequence 1, Appli
C	45	40	3.2	3366	US-08-267-803B-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-253-691-3/c
Sequence 3, Application US/09253691
Patent No. 6124100
; Publication/Patent: TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; TITLE OF INVENTION: Using Trinucleotide Repeats Sequence
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; CURRENT FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
US-09-253-691-3

Query Match   4.2%;   Score 51.8; DB 3; Length 397;
Best Local Similarity 89%; Conservative 0; Pred. No. 2.5e-05;
Matches   89; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy   873 GCTGCTGCTGCTGCCGCCGCCTCGTTGTTGTGGCTGCCAACTGCCGTTGTGCCGTTGTTCT 932
              ||||||||||||||||||||| |||||||||| ||||| |||||||||| |||||||||||
Db   200 GCTGCCGCTGCTGCTGCTGCTGCCGTGCTGCTGCTGGCTGCCGGCCGTTGTGCCGTTGTTGT 141

Qy   933 GCTGTGAGAAGAAAAAGAGAGGATTTCGTATTCAATTTCAAAAGAAAATCTGAAAAGAGAGA 992
              ||||| | |||||| || |||||||| |||||| ||| |||||| || || ||||||||||
Db   140 GCTGTTGCTGCTGCTTTTTGCTGCTGCTGCTGAACACATTCAAAAGTGAAGTATATTAAAAAAACAA 81

Qy   993 CAAACAAAGAAAAGAAACTGAGACAGCAGAAAGTGGAAA 1023
              |||| || |||| ||| | |||| |||| |||||| ||
Db   80 AACTTAAAAGAATAAATATACACCCATGAGAAAA 50

RESULT 2
US-08-531-927B-3/c
Sequence 3, Application US/08531927B
Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491
; TITLE OF INVENTION: Disease Gene and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

```
RESULT 15
BE376590

LOCUS       BE376590                663 bp    mRNA    linear    EST 21-JUL-2000
DEFINITION  601226419F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3584645 5',
            mRNA sequence.
ACCESSION   BE376590
VERSION     BE376590.1  GI:9321955
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 663)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8743 row: f column: 06
            High quality sequence stop: 498.

FEATURES             Location/Qualifiers
     source          1..663
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3584645"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH."

BASE COUNT      143 a    193 c    170 g    157 t
ORIGIN
```

```
Query Match    14.8%;  Score 184;  DB 10;  Length 663;
Best Local Similarity   73.4%;  Pred. No. 3.4e-32;
Matches   235;  Conservative   0;  Mismatches  85;  Indels   0;  Gaps   0;

Qy    115 AGGTTCTGGGTCTCGGTAAATGAAGTCATAGAAGGCCCCAGAAATGCAACAAGTCCTGAAGGG 174
          || ||||||||||| ||||| |||| ||||||| |||||||||| |||||||| |||
Db     91 AGCTTCCGGATCACATTATCAGATCATAGAAGGTCCTCAGAAATGTAACAGTCCTTAAAGGA 150

Qy    175 CTCCCAAGGCTCGCTTCAACTGCACCGTCTCCCAGGCGCTGGAAGCTCATCATGTTGGGCTCT 234
          ||| || |||||||||||| |||||| ||| ||||||| ||||| ||||||| |||| |||
Db    151 CTTCAGAGGCTCACTTCAACTGCACCACCGTGACTCACGCGTCTGGAAGCTTCTCATGTGGACTCT 210

Qy    235 CAGTGACATGGTGGTTGCTAAGCGTCAGGCCCATGGAGCCCATCATCACCAATGACCGCTT 294
          ||| ||||||||||||||| |||||||||| |||||||||||||| |||||||||||||||
Db    211 TAACCAAATGGTGGTTGCTGAGTCTCACCACCAAGGACCCATCACAACACAACCGCTT 270

Qy    295 CACCCTCAGAGGTACGACCGGGGGGGAACTTCACCTCGGGATGATCATCCACAATGT 354
          ||||| |||||||||||||| || ||
Db                                                                354
```

```
Db    296 CACCTATGCCAGTTACAAACAGCACACTGACAGCTCAATCTCGGAGTTGATCATCCATGATGT
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    355 GGAGCCCAGTGATTCGGGGAACATCAGATGCAGCGTCGCCTTCCAGAACAGTCGCATGGGATC 414
          |||||||| ||||| |||||||||||||||||||||||||||||||||||| |||||||||
Db    356 GCAGCCCAGTGACTCGGGATCCGTGCAAATGCAGCCATGGGTTTGGATC 415

Qy    415 TGCTTACCTTACCGTCCAAG 434
          |||||| ||| |||||||||
Db    416 TGCCTTCCTCTCAGTGCNAG 435
```

```
Db    271 CACCTATGCCAGTTACAAACAGCACACTGACAGCTCAATCTCGGAGTTGATCATCCATGATGT 330
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    355 GGAGCCCAGTGATTCGGGGAACATCAGATGCAGCGTCGCCTTCCAGAACAGTCGCATGGGATC 414
          |||||||| ||||| |||||||||||||||||||||||||||||||||||| |||||||||
Db    331 GCAGCCCAGTGACTCGGGATCCGTGCAAATGCAGCCATGGGTTTGGATC 390

Qy    415 TGCTTACCTTACCGTCCAAG 434
          |||||| ||| |||||||||
Db    391 TGCCTTCCTCTCAGTGCAAG 410
```

```
Search completed: April 29, 2003, 09:24:22
Job time : 2018.06 secs
```

```
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGAGAGAGAAGATCGCTCTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 5.0 and subtraction to
Rot = 20.0. Second strand cDNA was prepared with the primer adapter
of sequence[5'
GAGAGAGAGATTCTGAGTTTAATTAAATTAATCCCCCCCCCCC 3'], cDNA was cleaved
with XhoI and SstI. Cloning sites; 5' end: XhoI; 3' end: SstI.
Host: SOLR.
FEATURES             Location/Qualifiers
     source          1..1552
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:2010003D20"
                     /db_xref="MGD:MGI:1904786"
                     /clone="2010003D20"
                     /db_xref="taxon:10090"
                     /sex="male"
                     /tissue_type="small intestine"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     CDS             73..885
                     /note="data source:SPTR, source key:Q9NSI5, evidence:ISS
                     homolog to IGSF5 PROTEIN (FRAGMENT)
                     putative"
                     /codon_start=1
                     /protein_id="BAB25436.1"
                     /db_xref="GI:12842010"
                     /db_xref="MGD:MGI:1919308"
                     /translation="MEGSWRDVLAVLVILAQLTASGSSYQIIEGPQNTVLKDSEAHF
                     NCTVTHGWKLLMWTLNQMVVLSLTTQGPIITNNRFTYASYNSTDSFISELIIHDVQPS
                     DSGSVQCSLQNSHGPGSAFLSVQDSIGEGPALPTWAIILLAVAFSLLILIVLIII
                     FCCCASRREKEESTYQNEIRKSANMRTNKADPETKLKSGKENYGYSSDEAKAAQTAS
                     LPPKSAEVSLPEKRSSSLPYQELNKHQPGPATHPRVSFDIASPQKVRNVTLV"
     polyA_signal    1534..1539
                     /note="putative"
     polyA_site      1552
                     /note="putative"
BASE COUNT      430 a    369 c    374 g    379 t
ORIGIN

Query Match             15.0%;  Score 185.6;  DB 11;  Length 1552;
  Best Local Similarity   73.8%;  Pred. No. 1.6e-32;
  Matches  236;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy        115 AGGTTCTGGGGTCTGGTAATGAAGTCATAGAAGGCCCCCAGAATGCAACAGTCCTGAAGG 174
              || ||| || ||  || ||||||||||| ||||||||  || |||||||| ||||  ||
Db        129 AGCTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAATGTTAACAGTCCTAAAGGA 188

Qy        175 CTCCCAGGCTGCTTCCGGTTCAACTGCACCGTGTTCCCAGGGGTGGAAGCTCATCATGTC 234
              || |||||| | || || || |||  ||||  || ||||| ||||||||| ||||| |
Db        189 CTCAGAGGCTCACTTCAACTGCACCGTGACCCATGGGGCTGGAAGCTTCTCATGTGGACTCT 248

Qy        235 CAGTGACATGGTGGTTCAAGCGTCAGGCCCATGGACGCCATCATCACCAATGACCGCTT 294
              ||||||| ||||| ||||| |||  ||||||||||  | |||||||||||||||||||
Db        249 TAACAAAATGGTGGTTGCTGAGTCTCACCACCAAGGACCCATCATCACCAACAACCGCTT 308

Qy        295 CACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGAGATGATCATCCACAAGT 354
              ||||||||||||| | || |||||||||||||||  | ||| |||||||||  |||||
Db        309 CACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATCAATGATGT 368

Qy        355 GGAGCCCAGTGATTCGGGGGAACATCAGATGCAGCTCCCTCAGGTCGGCTGCAATGGATC 414
              |||| ||||||| | |||| ||||| ||||||||||| ||||||||  ||||||||| |||
Db        369 GCAGCCCAGTGACTCGGGATCCGTGCAATGCAGCCTGCAGAACAGCCATGGGTTTGGAATC 428

Qy        415 TGCTTACCTTACCGTCCAAG 434
              |||||||||||||||||||
Db        429 TGCCTCCTCTCAGTGCAAG 448
```

```
RESULT 14
AI425363

LOCUS       AI425363                 486 bp    mRNA    linear   EST 15-MAR-2000
DEFINITION  mx91d09.y1 Soares mouse NML Mus musculus cDNA clone IMAGE:693713
            5', mRNA sequence.
ACCESSION   AI425363
VERSION     AI425363.1  GI:4271294
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 486)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            This read was primed with a Not I - oligo(dT) primer (5'
            TGTTACCAATCTGAAGTGGGAGGCGGCCGCGCGAAACTCTTTTTTTTTTTTTTTT 3');
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
            correct orientation)
            MGI:427273
            Seq primer: -40RP from Gibco
            High quality sequence stop: 459
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..486
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:693713"
                     /clone_lib="Soares mouse NML"
                     /tissue_type="Liver"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer (5'
                     TGTTACCAATCTGAAGTGGGAGGCGGCCGCGCGAAACTCTTTTTTTTTTTTTTTT 3');
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      119 a    133 c    126 g    107 t      1 others
ORIGIN

Query Match             14.9%;  Score 184.6;  DB 9;  Length 486;
  Best Local Similarity   73.4%;  Pred. No. 2.3e-32;
  Matches  235;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy        115 AGGTTCTGGGGTCTGGTAATGAAGTCATAGAAGGCCCCCAGAATGCAACAGTCCTGAAGG 174
              || ||| || ||  || ||||||||||| ||||||||  || |||||||| ||||  ||
Db        116 AGCTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAATGTTAACAGTCCTAAAGGA 175

Qy        175 CTCCCAGGCTGCTTCCGGTTCAACTGCACCGTGTTCCCAGGGGTGGAAGCTCATCATGTC 234
              || |||||| | || || || |||  ||||  || ||||| ||||||||| ||||| |
Db        176 CTCAGAGGCTCACTTCAACTGCACCGTGACCCATGGGGCTGGAAGCTTCTCATGTGGACTCT 235

Qy        235 CAGTGACATGGTGGTTCAAGCGTCAGGCCCATGGACGCCATCATCACCAATGACCGCTT 294
              ||||||| ||||| ||||| |||  ||||||||||  | |||||||||||||||||||
Db        236 TAACAAAATGGTGGTTGCTGAGTCTCACCACCAAGGACCCATCATCACCAACAACCGCTT 295

Qy        295 CACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGAGATGATCATCCACAAGT 354
              |||||||| | ||| |  |  ||||||||||||  | ||| |||||||||  ||||
```

```
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM10260 row: n   column: 06
          High quality sequence stop: 628.

FEATURES             Location/Qualifiers
     source          1..931
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4459973"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     library constructed by Life Technologies. Investigator
                     providing samples: Gilbert, Smith, NIH."
BASE COUNT      233 a    277 c    242 g    179 t
ORIGIN
        1 ..931
```

```
Query Match        15.0%;    Score 185.6;   DB 12;   Length 931;
Best Local Similarity   73.8%;    Pred. No. 1.5e-32;
Matches   236;  Conservative    0;  Mismatches  84;  Indels   0;  Gaps   0;

Qy    115 AGGTTCTYGGGTCTGGTRAATGAAGTCATAGAAGAGGACCCCCAGANATGCAACAGTGCTCGAAGGG 174
          ||||||| ||||||| |||||||||||||||||||||||| ||  ||||||||||||||||| |||||
Db    157 AGCTTCCGGATCCAGTTATCAGATCATAGTAGGTCCTCCAGAAGGTCCTCAGAATGTAACAGTGCTCTAAGGA 216

Qy    175 CTCCCAGGCTCGGCTTCAACTGCACCGTCTCCCAGGGGTGGAGCTCATCAGTGGGCTCT 234
          ||| ||||||||| |||||||||||||| |||||||||||||||||||||||||||||||
Db    217 CTCAGAGGCTCACTCAACTGCACCGTGGAAGCTTCATGTGGACTCT 276

Qy    235 CAGTGACATGGTGTGCTAAGCGTCAGGCCCATGGGAGCCCATCATCACCANTGACCGCTT 294
          ||||||  |||||| |||||| ||  |||||||||||||||||||||||||| ||||||||
Db    277 TAACCAAATGGTGGTGGTGCTGAGTCTCCACCACCCAAGGACCATCATCACCACCAACCGCTT 336

Qy    295 CACCTCYCAGAGGTAGGACCCAGGGGCGGGAACTTCACCTCGAGATCAATCATCACAATGT 354
          ||||| |||||||| |||| |||| |||||||| |||||||||||| |||||||||||||||
Db    337 CACCTATGCCAGTTACAACAGACACTGACAGCTTCATCTCGGAGTTGATCATCCATGATGT 396

Qy    355 GGAGCCCAGTGATTCGGGGAACATCAGATGAGAGCCTCCAGAACAGTGCGCCTGCATGGATC 414
          | ||||||||||| |||||||| |||| || ||||| ||||| ||||||||| ||||||||
Db    397 GCAGCCCAGTGACTCGGGGATCCGTGCAAATGCAGCCCATGGGTTGGGATC 456

Qy    415 TGCTTACCTTACGGTCCAAG 434
          | |||| ||||| |||||||
Db    457 TCCTTCCTCTCAGTGCAAG 476
```

```
RESULT 13
LOCUS       AK008060     1552 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male small intestine cDNA, RIKEN full-length
            enriched library, clone:2010003D20:homolog to IGSF5 PROTEIN
            (FRAGMENT), full insert sequence.
ACCESSION   AK008060
VERSION     AK008060.1  GI:12842009
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
REFERENCE   2
  PUBMED    10349636
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
  PUBMED    11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
  PUBMED    11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C.A., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
  PUBMED    11217851
REFERENCE   5  (bases 1 to 1552)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J.,
            Kato,H., Kawai,J., Kojima,Y., Kondo,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Nunazaki,C., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Direct Submission
            Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
```

Left column

```
MGI:127273
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 454.
```

FEATURES   Location/Qualifiers
source   1..471
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone="IMAGE:693713"
   /clone_lib="Soares mouse NML"
   /tissue_type="Liver"
   /lab_host="DH10B"
   /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
   polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
   was primed with a Not I - oligo(dT) primer [5'
   TGTTACCAATCTGAAGTGGGACGGCGGCGCGGAATCTTTTTTTTTTTTTTTTTT 3'];
   double-stranded cDNA was ligated to Eco RI adaptors
   (Pharmacia), digested with Not I and cloned into the Not I
   and Eco RI sites of the modified pT7T3 vector. Library
   constructed and normalized by Bento Soares and M.Fatima
   Bonaldo."

BASE COUNT   118 a   126 c   122 g   105 t
ORIGIN

Query Match   15.0%;   Score 185.6;   DB 9;   Length 471;
Best Local Similarity   73.8%;   Pred. No. 1.4e-32;
Matches 236;   Conservative 0;   Mismatches 84;   Indels 0;   Gaps 0;

```
Qy   115 AGGTTCTGGGTCTGGGTAATGAAGTCATATAGAAGGCCCCCAGAAATGCAACAGTCCTGAAGG 174
             |||||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   116 AGCTTCCGGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAAATGTAACAGTCCTAAAGGA 175

Qy   175 CTCCCAGGCTCGCGTCTCCCAGGGCTGGGAGGCTCATCATCATGTGGGGCTCT 234
             ||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   176 CTCAGAGGCCTCACTTCAACTGCACCGTGACTCACCGGCTTCTCATGTGGACTCT 235

Qy   235 CAGTGACACATGGTGGTGTGCTAAGCGTCAGGCCCATGGAGCCCATCATCATGACCGGCTT 294
             ||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db   236 TAACCAAATGGTTGGTTGCTCACCACCAGGACCCATCATCACCAACACCGCTT 295

Qy   295 CACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGGAGATGATCATCCACAATGT 354
             ||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||||
Db   296 CACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATCCATGATGT 355

Qy   355 GGAGCCCAGTGATTCGGGGGAACATCAGATGCAGCCTCCAGAACAGTCGCCTGCAATGGATC 414
             ||||||||||||||| |||| ||||||||||||||||||||||||||||||| ||||||
Db   356 GCAGCCCAGTGACTCGGGATCCGTGCAATGCACCCTGCAGAACAGCAGCCATGGGTTTGGATC 415

Qy   415 TGCTTACCTTACCGTCCAAG 434
             |||||||||||||| |||||
Db   416 TGCCTTCCTCTCAGTGCAAG 435
```

```
RESULT 11
LOCUS       AW990468   611 bp   mRNA   linear   EST 02-JUN-2000
DEFINITION  uf37d05.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA
            clone IMAGE:1513545 5' similar to TR:O60962 O60962 L549.1. ;, mRNA
            sequence.
ACCESSION   AW990468
VERSION     AW990468.1  GI:8186027
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 611)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
```

Right column

```
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:940397
Seq primer: -40RP from Gibco
High quality sequence stop: 474.
```

FEATURES   Location/Qualifiers
source   1..611
   /organism="Mus musculus"
   /db_xref="taxon:10090"
   /clone="IMAGE:1513545"
   /clone_lib="Soares_mammary_gland_NMLMG"
   /sex="female (lactating)"
   /tissue_type="mammary gland"
   /lab_host="DH10B"
   /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
   polylinker; 1st strand cDNA was prepared from mammary
   gland tissue from a lactating female, and was then primed
   with a Not I - oligo(dT) primer. Double-stranded cDNA was
   ligated to Eco RI adaptors (Pharmacia), digested with Not
   I and cloned into the Not I and Eco RI sites of the
   modified pT7T3 vector. Library is normalized. Library
   was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT   152 a   155 c   156 g   148 t
ORIGIN

Query Match   15.0%;   Score 185.6;   DB 10;   Length 611;
Best Local Similarity   73.8%;   Pred. No. 1.4e-32;
Matches 236;   Conservative 0;   Mismatches 84;   Indels 0;   Gaps 0;

```
Qy   115 AGGTTCTGGGTCTGGGTAATGAAGTCATATAGAAGGCCCCCAGAAATGCAACAGTCCTGAAGG 174
             |||||| |||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   125 AGCTTCCGGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAAATGTAACAGTCCTAAAGGA 184

Qy   175 CTCCCAGGCTCGCGTCTCCCAGGGCTGGGAGGCTCATCATCATGTGGGGCTCT 234
             ||| |||||||||||||||||||||||||||||||||||||||||| ||||
Db   185 CTCAGAGGCCTCACTTCAACTGCACCGTGACTCACCGGCTTCTCATGTGGACTCT 244

Qy   235 CAGTGACACATGGTGGTGTGCTAAGCGTCAGGCCCATGGAGCCCATCATCATGACCGGCTT 294
             ||||| ||||| ||| ||||||||||||||||||||||||||||||||||||||||||||
Db   245 TAACCAAATGGTTGGTTGCTCACCACCAGGACCCATCATCACCAACACCGCTT 304

Qy   295 CACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGGAGATGATCATCCACAATGT 354
             ||||||||| |||||||||||||||||||||||||||||||| ||||| ||||||||||
Db   305 CACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATCCATGATGT 364

Qy   355 GGAGCCCAGTGATTCGGGGGAACATCAGATGCAGCCTCCAGAACAGTCGCCTGCAATGGATC 414
             ||||||||||||||| |||| ||||||||||||||||||||||||||||||| ||||||
Db   365 GCAGCCCAGTGACTCGGGATCCGTGCAATGCACCCTGCAGAACAGCAGCCATGGGTTTGGATC 424

Qy   415 TGCTTACCTTACCGTCCAAG 434
             |||||||||||||| |||||
Db   425 TGCCTTCCTCTCAGTGCAAG 444
```

```
RESULT 12
BG173684
LOCUS       BG173684   931 bp   mRNA   linear   EST 06-FEB-2001
DEFINITION  602336823F1 NCI_CGAP_Mam1 Mus musculus cDNA clone
            IMAGE:4459973 5', mRNA sequence.
ACCESSION   BG173684
VERSION     BG173684.1  GI:12680387
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 931)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
```

RESULT 9

LOCUS   AV601192   524 bp   mRNA   linear   EST 27-NOV-2001
DEFINITION AV601192 Bos taurus kidney fetus Bos taurus cDNA clone E1KI002H04
            5', mRNA sequence.
ACCESSION   AV601192
VERSION   AV601192.1 GI:9723510
KEYWORDS   EST.
SOURCE   cow.
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 524)
  AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso,H.
            and Sugimoto,Y.
  TITLE   Establishment of a high throughput EST sequencing system using
            poly(A) tail-removed cDNA libraries and determination of 36,000
            bovine ESTs
  JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
  MEDLINE   21570554
  COMMENT   Contact: Yoshikazu Sugimoto
            Animal Genetics Division
            Shirakawa Institute of Animal Genetics
            Odakura, Nishigo, Nishi-Shirakawa, Fukushima 961-8061, Japan
            Tel: 81-248-25-5641
            Fax: 81-248-25-5725
            Email: kazusugi@cocoa.ocn.ne.jp
            Single pass sequencing.

FEATURES   Location/Qualifiers
     source   1..524
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone="E1KI002H04"
                     /clone_lib="Bos taurus kidney fetus"
                     /tissue_type="kidney"
                     /dev_stage="fetus"
                     /note="Vector: pZL1; Site_1; SalI; Site_2: NotI; Poly A
                     was deleted from a NotI site"
BASE COUNT   120 a   135 c   168 g   96 t   5 others
ORIGIN

Query Match   15.3%; Score 190; DB 10; Length 524;
Best Local Similarity   76.1%; Pred. No. 1.3e-33;
Matches 232; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy   233 CTCAGTGACATGGTGGTGCTAAGCGTCAGGCGCCCATGGAGCCCATCACCAATGACCGGC 292
              ||| |||| |||| |||||| |||||||| |||||||||| ||||||||||| ||||||
Db   280 GGCTCCGGAGGCTTCGCTTCAACTGCACCATCTCCCAGGGGCTGGAAGCTTCATTCATGTG 339

Qy   173 GGCTCCCAGGCCTCGGCTTCAACTGCACGTCCTCCCAGGGGCTGGAAGCTTCATTCATGTG 232
              |||||||||||||| || ||||||||||||||| |||||| |||| ||||||||||||||
Db   220 GCAGCCGTGTGGATCCAGCAGTGAAAATCATAGAGAGGGTTCCCAAGAATGTCACAGGCCCTGAAG 279

Qy   113 GTAGGTTCTGGGTCTCGGTAATGAAGTCATAGAAGGCCCCCCAGAAATGCAACAGTGCCTGAAG 172
              | || |||| ||| |||| || || |||||| |||| |||||| |||||| || ||||| |||
Db   160 CTGAGAGGCACAGTGGTGGTCGTGAGCATGACACTTAATGACCATCATCACCAGTGACCGC 399

Qy   293 TTCACCTCTCAGAGGTACGACCAGGCGCGGGAGAACTTCACCTCCGGAGATGATCATCCACAAT 352
              |||||| || ||||||||| ||||||| |||||||||||||||||||| ||| ||||| |||
Db   400 TTCACTTCGGCAAGTACCAAGAGAGGGCCGGGAACTTCATCTNTNGAGATGAATAATTCATGAC 459

Qy   353 GTGGAGCCCAGTGATTCGGGAACAATCAGATGCAGCCTCCAGAGACAGTGCCCTTCCATGGA 412
              |||| ||||| ||| |||||||| ||| || |||||| |||||||||| || |||| ||||
Db   460 GTGCAACTCACCGATCCCGGCGAAGTNAAAATGCAGCCTTCANAACAGCAATCGGGATGGA 519

Qy   413 TCTGC 417
              |||||
Db   520 GCCGC 524

RESULT 10

LOCUS   AA265274   471 bp   mRNA   linear   EST 20-MAR-1997
DEFINITION AA265274 mx91d09.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:693713
            5', mRNA sequence.
ACCESSION   AA265274
VERSION   AA265274.1 GI:1901370
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 471)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE   The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
  COMMENT   Contact: Marra M/Mouse EST Project
            Washington University School of Medicine
            WashU-HHMI Mouse EST Project
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.

This clone was obtained from a polyA-deleted cDNA library.

FEATURES   Location/Qualifiers
     source   1..524
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-112N07.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT   168 a   159 c   143 g   207 t
ORIGIN

Query Match   15.4%; Score 190.8; DB 17; Length 677;
Best Local Similarity   94.3%; Pred. No. 9e-34;
Matches 198; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy   539 TCCTGGGAGCTCGGGTCTCGTCAGCCATTCATCAAGCTATTATATTNTTGTTCCGGAGCCCAGC 598
              || ||||||| |||||||||||| || || |||||||||| ||||| ||||||||||||||||
Db   72 TCCGCGGAGCGTCGGGTCTTCCTGGTCCACCCATTCAAGCTATTATATTTGTTCCGGAGCCCAGC 131

Qy   599 GACCTTCAAAGTGCAGTGAGCATCCTTGGCTCTGACCCCACAGAGCAATGGGACTTTGACT 658
              ||||||||||||||||||||||||| ||||||||||||||||||| ||| |||||||||||
Db   132 GACCTTCAAAGTGCAGTGAGCATCCTTGGCTCTGACCCCACAGAGCAATGGGACTTTGACT 191

Qy   659 TGCGTGGCTACCTGGAAGAGAGGCCTGAAGGCCCCGGCAACAGTCTTCGCAACGTGTAAAATTCTCACTGTG 718
              |||||||| |||||||||||| || ||| ||||||| || |||||||| |||| ||||||||||||||||||
Db   192 TGCGTGGCTACCTGGAAGAGAGGCCCCACAAGGTCTGCAACTGTAAAATCTCACTGTG 251

Qy   719 AATCGGTGTCCCCAAGACACTGGAGGTGGT 748
              |||||||||||||||||||| ||||| |||
Db   252 AATCGGCGTCCCCAAGGTAAGTGAAGATGT 281

USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@anri.barc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
vO.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 27 row: H column: 19
Seq primer: ATTTAGGTGACACTATAG.

FEATURES   Location/Qualifiers
source   1..517
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="BARC 5BOv"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled mRNA isolated from mammary
                     tissues at eight physiological, developmental, and disease
                     states."

BASE COUNT   117 a   133 c   165 g   102 t
ORIGIN

Query Match
Best Local Similarity   15.9%;   Score 197.4;   DB 10;   Length 517;
Matches 243;   Conservative 0;   Pred. No. 2.5e-35;   Mismatches 76;   Indels 0;   Gaps 0;

Qy   113 GTAGGTTCTGGGGTCGTTGGTAATGAAGTCATAGAAGAGGCCCCCAGAGAATGCAACAGTCCTGAAG 172
          ||| || || | || ||| |||| ||||| || |||| || ||| |||||| ||||||
Db   199 GCAGCCTGTGGGATCCAGCAGTGAAATCATPAGAGGGTTCCCAAGAAATGTCACAGCCCTGAAG 258
Qy   173 GGCTCCCAGGCTCGCTTCTTCCCAGGCTGTCCGGAAGCTCATCATGTGGGCT 232
          ||||| ||||||||||| || | || | ||||||||| | ||||||||||
Db   259 GGCTCGGAGGCTCGCTTCAACTGCACCACTCTCCCAGGGCTGGGAAGCTCGTCACATGTGGGCT 318
Qy   233 CTCAGTGACATGGGTTGGCTAAGGCTCCCCCATGGGAGCCCATCATCAACAATGACCGC 292
          || || || || |||||| ||||| ||||||| ||||||||||| || ||||||||
Db   319 CTGAGAGGCAACAGTGGGTCGTCTGACCATGACACCTPAATGAGACCATCATCACCAGTGACCGC 378
Qy   293 TTCACCTCTCAGAGGTACCGACCCAGGCGCGGGAACTTCACCTCACGTCGGAGGATGATCATCCACAAT 352
          ||||| | | || | ||| || | || |||||||||| | ||| | ||||| |||| | ||||| |
Db   379 TTCACTTCGGCAAGCTACCAAAGAGAGGGCGGGAACTTCAACCTCTCGAGATGATAATTCATGAC 438
Qy   353 GTGGGAGCCCAGTGAGTGATTCGGGGGACATACAGATGGCAGCCTCCAGCGAACAGTGCCGCAIGGA 412
          ||| | | |||| | | || ||| | || ||| || ||||||| ||| ||||||| | ||| |||
Db   439 GTGCAACTCAGCCGANGCCCGGCAAGTCAAATGCAGCCTTCAGACAACAGCAATCGGGATGGA 498
Qy   413 TCTCGCTTACCTTACCGTCC 431
          | ||| || ||||| |||||
Db   499 GCCGCCTTCCTTTCTGTTC 517

RESULT 7
LOCUS   AG069679   622 bp   DNA   linear   GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-060F12.F, genomic survey sequence.
ACCESSION   AG069679
VERSION   AG069679.1 GI:16621481
KEYWORDS   GSS.
SOURCE   Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-060F12.F.
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   BAC end sequences of Library PTB
JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 622)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
            1-7-22 Suehiro-chou Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT   Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
            Sequencing: -21M13
            LIBRARY
            Vector   : pKS145
            R. Site 1   : SacI
            R. Site 2   : SacI

FEATURES   Location/Qualifiers
source   1..622
                     /organism="Pan troglodytes"
                     /db_xref="taxon:9598"
                     /clone="PTB-060F12.F"
                     /sex="male"
                     /cell_type="lymphoblast"
                     /clone_lib="PTB 060F12.F"

BASE COUNT   165 a   153 c   128 g   176 t
ORIGIN

Query Match
Best Local Similarity   15.4%;   Score 190.8;   DB 17;   Length 622;
Matches 198;   Conservative 0;   Pred. No. 8.9e-34;   Mismatches 14;   Indels 0;   Gaps 0;

Qy   539 TCCTGGGAGGCTCGGTCTCCTCGGTGTCAGCCATTCAAAGCTATTATTTGTTCCGGAGCCCAGC 598
          || | ||||||||||| | | ||| ||||||||||||||||||||||||||||||||||||||
Db   72 TCCGCGGGAGGCTCGGTCGTTCCTGGTCAGCCATTCAAAGCTATTATTTGTTCCGGAGCCCAGC 131
Qy   599 GACCTTCAAAGTGCAGTGAGCATCCTGCGTCTCACCCCACAGAGCAATGGGACTTTGACT 658
          |||||||||||||||||||||| ||||| | ||| |||||||||||||||||||||||||
Db   132 GACCTTCAAAGTGCAGTGAGCAATCCTGGCTCTGACCCCACAGAGCAATGGGACTTTGACT 191
Qy   659 TGCGTGGCTACCTGGAAGAGAGCCTGAAGAGCCCCGCCAAGTCTGCAACTGTAAATCTCACTGTG 718
          ||| ||||||||||||||||||||||||||||||| | |||||||||||||||||||||||||||
Db   192 TGCCTGGCTACCTGGAAGAGAGCCTGAAGAGCCCCACAAGTCTGCAACTGTAAATCTCACTGTG 251
Qy   719 ATTCGGGTGTCCCCAAGACACACTGGAGGTGGT 748
          |||||| |||||||||| | | ||||| ||
Db   252 ATTCGGCGGTCCCCAAAGGTAAGTGAAGAGATGT 281

RESULT 8
LOCUS   AG107877   677 bp   DNA   linear   GSS 03-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-112N07.F, genomic survey sequence.
ACCESSION   AG107877
VERSION   AG107877.1 GI:16728395
KEYWORDS   GSS.
SOURCE   Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
            BAC Library clone:PTB-112N07.F.
ORGANISM   Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1 (bases 1 to 677)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Unpublished
JOURNAL
REFERENCE   2 (bases 1 to 677)
AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE   Direct Submission
JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

RESULT 5 block

```
             primed with a primer (5'
             GAGAGAGAGCGGCGCGCCAACTCGAGTTTTTTTTTTTTTTTVN 3'], cDNA was
             prepared by using trehalose thermo-activated reverse
             transcriptase and subsequently enriched for full-length by
             cap-trapper. Second strand cDNA was prepared with the
             primer adapter of sequence [5'
             GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
             was cleaved with BamHI and XhoI. Vector: a modified
             pBluescript KS(+) after bulk excision from lambda FLC I."
```

```
BASE COUNT    154 a   167 c   161 g   149 t
ORIGIN
```

```
Query Match    20.9%;  Score 259;  DB 10;  Length 631;
Best Local Similarity  72.2%;  Pred. No. 1.4e-49;
Matches  337;  Conservative  0;  Mismatches 130;  Indels  0;  Gaps  0;
```

```
Qy   115 AGGTTCTGGGTCTGTAATGAAGTCATAGAAGGCCCCACAGTCCTGAAGG 174
         ||| ||| ||| || ||  |||| ||||||||||  |||||| ||||
Db   162 AGCTTCCGGATCCAGTTATCAGATCATAGAAGGTTCCTCAGTCCTAAGGA 221

Qy   175 CTCCCAGGCGCTCGCGTTCAACTGCACCGTCTCCCAGGCGTGGGAAGCTCATCACATGTGGGGCTCT 234
         | ||| || || |  || |||| ||||||    || ||||| ||||| || |||||
Db   222 CTCAGAGGCTCACTTCACCTGCACCGTGACTCACGGCGTGGAAGCTTCCATGTGGGACTCT 281

Qy   235 CAGTGACAATGGGTGGTGCTAAGCGTCAGCCCCATGGAGCCCATCACCAATGACCGCTT 294
         |||| ||| ||||||||||| ||  ||  |||| ||| |||||||| |||||| |||||
Db   282 TAACCAAATAGGGTGGTGCTGAGTCTCCACCACCCAAGGACCCATCATCACCAACAACCGCTT 341

Qy   295 CACCTCTCAGAGGTACGACCAGGGGGGAACTTCACCTCGGACGATGACATCCACAATGT 354
         |||| |||||||| |  ||||| |||| |||||  |||| || ||||||| |||||
Db   342 CACCTATGCCAGTTACAACAGCAGCTTCATCTCGGAGTTGATCATCCATGATGT 401

Qy   355 GGAGCCCAGTGATTCGGGAACATCAGATGCAACAGCAGTCGCCCTGCAATGGATC 414
         ||||||| |||| |||| || ||||| ||| |||||||  |||||| |||||||
Db   402 GCAGCCCAGTGACTCGGGATCCGTGCAATGCACGGCAGTGGGTTTGGATC 461

Qy   415 TGCTTACCTTACCGTCCAAGTTATGGGAGAGCGTGTTCATTCCCAGTGTTAATCTTGTAGT 474
         |||  |||||||||  |||||  | | |||  ||| ||||| ||| |||  |||| ||
Db   462 TGCCTTCTCTCAGTGACCAAGTCCATGGGGACCCTGAACACAACCTTATAGT 521

Qy   475 CGCTGAGAATGAACCTTGTGAAGTTACTTGTTGTCTACCCTACACCGGGGTCCCGGA 534
         | ||  ||||| || ||||| || |||  | |||  |||| |||| || |||||||
Db   522 CACTGAGGGTGAACCCTGTAATGGACTTGCTATGCCCTGGGTGGACTCACTCCCGGA 581

Qy   535 TATTTCCTGGGGAGCTCGGTGTCTCCTGGTGCAGCCAATTCAAGCTATTATT 581
         ||||||| |||||||| ||| ||||| |||| || |||| | |||| ||
Db   582 TATTTCCTGGGGAGCTTGAGGTTCCCGTAAGCCATTCAGAGTTACAATT 628
```

```
FEATURES             Location/Qualifiers
     source          1..527
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
```

```
RESULT 5
LOCUS        BE032610       527 bp    mRNA    linear   EST 09-JUL-2000
DEFINITION   132035 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION    BE032610
VERSION      BE032610.1  GI:8327619
KEYWORDS     EST.
SOURCE       pig.
ORGANISM     Sus scrofa
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE    1  (bases 1 to 527)
AUTHORS      Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
             Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
             and Keele,J.W.
TITLE        Design and use of two pooled tissue normalized cDNA libraries for
             EST discovery in swine
JOURNAL      Unpublished (2000)
COMMENT      Contact: Smith TPL
             USDA, ARS, US Meat Animal Research Center
             PO Box 166, Clay Center, NE 68933-0166, USA
             Tel: 402 762 4366
             Fax: 402 762 4390
             Email: smith@email.marc.usda.gov
```

RESULT 6 block

```
             Single pass sequencing. Bases called and alt_trimmed with phred
             v0.980904.e. Vector identified by cross_match with the -minscore 18
             and -minmatch 12 options.
             PCR Primers
             FORWARD: AGGAAACAGCTATGACCAT
             BACKWARD: GTTTTCCCAGTCACGACG
             Plate: 63   row: E  column: 13
             Seq primer: ATTTAGGTGACACTATAG.
```

```
FEATURES             Location/Qualifiers
     source          1..527
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
```

```
BASE COUNT    105 a   147 c   176 g    99 t
ORIGIN
```

```
Query Match    19.9%;  Score 246.2;  DB 10;  Length 527;
Best Local Similarity  79.0%;  Pred. No. 1.2e-46;
Matches  293;  Conservative  0;  Mismatches 78;  Indels  0;  Gaps  0;
```

```
Qy   113 GTAGGTTCTGGGGTCTGGTGTTAATGAAGTCATAGAAGGGCCCCACAGAATGCAACAGTCCTGAAG 172
         || ||| |||||| |||||| || |||||||| | |||||| |  |||| | ||| |||| |||||
Db   157 GCAGCTTGTGGATCCGGCAATGAAATCTTAGAAGGGTCCAGGGTCCAGGAAACGTCACGGGTGAAG 216

Qy   173 GGCTCCCAGGCGCTCGCTTCAACTGCACCGTCTCCCAGGCGTGGGAAGCTCATCACATGTGGGGCT 232
         |||| |||||| || |||| |||||||||| | |||||| || |||||| | |||||||| |||||
Db   217 GGCTCGGAGGCTCGGCTTCAAACTGCACCATCTCCCAGGGCTGAAGCTTCATCAGTGTGGGGCT 276

Qy   233 CTCAGTGACAATGGTGGTGGTGCTAAGCGTCAGGCCCCATGGAGCCCATCATCACCAAATGACCGC 292
         ||||||||| ||||| ||||| ||| |||  | ||||||||| |||||| ||||| || | ||| |
Db   277 CTGAAAAGGCACGGTGGGTTCTGAGCGTCACGGCGCCCATCATCACCACGACCGA 336

Qy   293 TTCACCTCTCAGAGGTACGACCAGGGGGCGGGAACTTCACCTCGGACGATGACATCATCACAAT 352
         ||||||| | ||||| | |||||||| ||||| ||||||||| |||| ||||| || ||||
Db   337 TTCACGTCTGCAAGTTATGAAGAGGGGCGGTAAGCTCATTCCTCGAGATGAGATCATCCATGAC 396

Qy   353 GTGGAGCCCAGTGATTGGGGAACATCAGATGCAACAGTCGCCTGCAGCGTCGCTGCATGGA 412
         |||||||||||| |||| ||| ||||| ||| || || || || |||||| ||| |||||
Db   397 GTGCAGCCTCAGTGATGCCGGCAGGTCAAATGCAGGAGCCTCCAGAACAGTGACCGGGATGGA 456

Qy   413 TCTGCCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTTCATTCCCAGTGTTAATCTTGTA 472
         ||| |||| |||||||| |||||| | ||||||| ||| |||| ||| |||  |||| ||
Db   457 TCCGCCTTTCCTTTCCGTCCAAGTCATGGGGTGCTGCTCATTCCCGAGGCAGCCTCGTA 516

Qy   473 GTCGCCTGAGAA 483
         ||| ||||  |
Db   517 GTCCTTGCGGA 527
```

```
RESULT 6
LOCUS        BE476432       517 bp    mRNA    linear   EST 28-AUG-2000
DEFINITION   159670 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION    BE476432
VERSION      BE476432.1  GI:9595965
KEYWORDS     EST.
SOURCE       cow.
ORGANISM     Bos taurus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
             Bovinae; Bos.
REFERENCE    1  (bases 1 to 517)
AUTHORS      Sonstegard,T.S., Capuco,A.V., Van Tassell,C.P., Ashwell,M.S. and
             Wells,K.D.
TITLE        Mapping of Expressed Sequence Tags from a normalized bovine mammary
             gland cDNA library
JOURNAL      Unpublished (2000)
COMMENT      Contact: Sonstegard TS
```

```
RESULT 3

LOCUS       BG740428                725 bp    mRNA    linear   EST 15-MAY-2001
DEFINITION  602638817F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4778789 5',
            mRNA sequence.
ACCESSION   BG740428
VERSION     BG740428.1  GI:14051081
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 725)
AUTHORS     NIH-MGC
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            http://image.llnl.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1635 row: b column: 06
            High quality sequence stop: 725.
FEATURES             Location/Qualifiers
     source          1..725
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4778789"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT      227 a    163 c    159 g    176 t
ORIGIN

Query Match       21.4%;  Score 265.8; DB 12; Length 725;
Best Local Similarity   99.3%;
Matches  267;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy    971 AAGAAAATCTGAAAAAAGAGAAGGAGACAAACAAAACTGAGACAGAAAAGTGGAAATGAAAAC 1030
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    377 AGGAAAATCTGAAAAAGAGAACAAACAAAACTGAGACAGAAAGTGGAAATGAAAAC 436

Qy   1031 TCCGGCTACAATTCAGATGAACAAAAAGACAACAGAAAACCGCTTGTCTCCCTTCCCAAATCC 1090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    437 TCCGGCTACAATTCAGATGAACAAAAAGACAACAGAAAACCGCTTGTCTCTCCCTTCCCAAATCC 496

Qy   1091 TGTGAATCCAGTGGATCCTGAACAAAAGACAGTAGCTGGTGGCCTTCTCTTCTTTTAATCTGGCCAGTCCT 1150
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    497 TGTGAATCCAGTGGATCCTGAACAAAAGACAGTAGCTGTGGCCTTCTCTTCTTTTAATCTGGCCAGTCCT 556

Qy   1151 GATCAACGTCCACCCAGGCCCAGGCCAAGTCATCCACAGGCCTTCTTTTAATCTGGCCAGTCCT 1210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    557 GATCAACGTCCACCCAGGCCCAGGCCAAGTCATCCACAGGCCTTCTTTTAATCTGGCCAGTCCT 616

Qy   1211 GAGAAGGTCAGTAATACAACTGTAGTATA 1239
          |||||||||||||||||||||||||||||
Db    617 GAGAAGGTCAGTAATACAACTGTAGTATA 645
```

```
Db  629  --GGGCGACTTTATGAGGGTCTGGAGGAGGCCTGAAGGCCCGCCAAGTCT
         |||| |||| |||  ||||  |||||| ||  |||| |||| ||| ||||||
Qy  655  GACTTGCGTGGCCTACCTGGAAGAGGCTGAGTGTCTCTGGAACTCA-CACACTGGGCAACGGGACCTT 697
         |||||| ||||| |||||||||||||||||||||||| ||| |||||
Db  686  GACTTGTGTGGCCAAGAGCTGAAGGACTTGCAGGCCAGCAGTCT 728
```

```
RESULT 4

LOCUS       BB663870                631 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB663870 RIKEN full-length enriched, 0 day neonate lung Mus
            musculus cDNA clone E030016M12 5', mRNA sequence.
ACCESSION   BB663870
VERSION     BB663870.1  GI:16497624
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 631)
AUTHORS     Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL     Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh'
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            Watahiki,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            , S., Kawai,J., Okazaki,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            Hayashizaki,Y.
            Computational Analysis of Full-length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..631
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="E030016M12"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     lung"
                     /tissue_type="lung"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
```

```
source          1..784
                /organism="Mus musculus"
                /strain="NMRI"
                /clone="IMAGE:491803"
                /db_xref="taxon:10090"
                /clone_lib="NCI_CGAP_Mam4"
                /tissue_type="tumor, gross tissue"
                /dev_stage="5 months"
                /lab_host="DH10B"
                /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                Library constructed by Life Technologies. Investigators
                providing samples: Lothar Hennighausen/Priscilla Furth,
                NIH Reference for transgenic model: Li et al., Cell Growth
                and Differentiation 7, 3-11 (1996)."

BASE COUNT      188 a    202 c    208 g    186 t
ORIGIN

Query Match        24.9%;  Score 308.2;  DB 12;  Length 784;
Best Local Similarity   70.8%;
Matches  437;  Conservative   0;  Mismatches 178;  Indels   2;  Gaps   2;
                        Pred. No. 5.4e-61;

Qy  115 AGGTTCTGGGTCTGGGTAAIGAAGTCATAGAAGGCCCCAGAAATGCAACAGTCCTGAAGGG 174
        | ||| ||| ||| ||||  |||| ||| |||| ||| ||||||| ||  || || |||
Db  149 AGCTTCCGGATCCAGTTATCAGATCATAGAAGGGTCCTCAGAAGTGTAACAGTCCTAAAGGA 208

Qy  175 CTCCCAGGCGTCGCTTCAACTGCACCGTGTCCCAGGGCTGGAGCGTCATCATGTGGGCTCT 234
        ||| ||||| | ||||||||||||||||| || |||||||| || | |||||||| ||||
Db  209 CTCAGAGGCTCACTTCAACTGCACCGTGACTCACGGCTGGAAGCTTCTCATGTGGACTCT 268

Qy  235 CAGTGACAATGGTGGTGCTAAGGCTGTCAGGCCCCATGGAGCCCATCATCACCAATGACCGCTT 294
        | ||| ||||||||||||||  | ||  |  | | ||  ||||||||||||||| | | ||||
Db  269 TAACCAAAATGGTGGTGCTCGAGTCTCACCACCAAGGACCCCATCATCACAACCGCTT 328

Qy  295 CACCTCTCAGAGGTACGACCCAGGGCGGGGAACTCACCTCGGAGATGATGATCATCACAAATGT 354
        ||||| | ||| | ||  || |   | | ||   ||||||||| | ||||| |  ||||| |||
Db  329 CACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATTCATGATGT 388

Qy  355 GGAGCCCAGTGAATCGGGGAACAATCAGATGACAGCCTCCAGGAACAGTCGCCCTGCATGGATC 414
        | | |||||||| | ||||| |   | | || ||||| ||| |||| |  |  ||| | |||||
Db  389 GCACCCCAGTGACTGCGGGATCCGTGCAATGCAGCAGAACAGCCGCCATGGGTTTGGATC 448

Qy  415 TGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTGTTCATTCCCAGTGTGTTAATCTTGTAGT 474
        |||  | ||| | ||| |||| ||||||||  | | || ||||| || | | | |  || ||||
Db  449 TGCCTTCCTTCTCAGTGCAAGTCATGGGAGACCCTGAACATTCCTAGCGAACAACTTATAGT 508

Qy  475 CGCTGAGAATGAAGACCTTGTGAAGTTACTTGTGTCTACCCTCACACTGGACCCGGCGCTT 534
        | |||||  ||||  ||| || | | || || | | || ||||||| ||||||  ||  |
Db  509 CACTGAGGGTGAACCCTGTAATGTGACTTGCATGCCCTGGCTGGACCCTCACTCCCGGA 568

Qy  535 TATTTCCTGGGAGCTCGGGTCTCCTGTGGCTAGCCATTCAAGCTATTATTTTGTTCCGGAGCC 594
        |||||||||||||| | ||| |   | | |  |||||| ||| || | | ||  ||| |||||
Db  569 TATTTCCTGGGAGCTTGAGGTTCCCGTTAAGCCATTCGAGTTAACAAACCTTATAGT 628

Qy  595 CAGGGACCTTCAAAGTGCAGTGAGCATCCTGGACCCCACAGAGCAATGGGGACTTT 654
        | ||| |||| | ||||| |||| | | | |||||| |||| | |  |||| ||||
Db  629 GGGCAACTTTATGAGGGTCTTGAGGTGTCTGGACCTCACACA-TGGCACGGACCTT 687

Qy  655 GACTTGCGTGGGCTGTACCTGGAAGGCCGGCAAGTGCTGCAACTGAAATCTCAC 714
        |||||| ||||| | | || || | |  | || || || ||||| | |  |||
Db  688 GACTTGTGTGGGCAGAGACTTGAAGGACTCGCAGCAGTCCTTAACTGTGTCAA-CTGAC 746

Qy  715 TGTGTCAATTCGGTGTCCCC 731
        ||||  | || ||  |  ||
Db  747 TGTGGTTCAGCTTCCAC 763
```

RESULT 2
BF784177

```
LOCUS       BF784177     794 bp    mRNA    linear   EST 12-JAN-2001
DEFINITION  602100039F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4236489
            5', mRNA sequence.
ACCESSION   BF784177
VERSION     BF784177.1  GI:12089213
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 794)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
  TITLE     Unpublished (1999)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9844 row: n column: 10
            High quality sequence stop: 694.
FEATURES             Location/Qualifiers
     source          1..794
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4236489"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library.|"

BASE COUNT      186 a    214 c    215 g    179 t
ORIGIN

Query Match        22.5%;  Score 278.6;  DB 12;  Length 794;
Best Local Similarity   69.6%;
Matches  406;  Conservative   0;  Mismatches 174;  Indels   3;  Gaps   2;
                        Pred. No. 4e-54;

Qy  115 AGGTTCTGGGTCTGGGTAAIGAAGTCATAGAAGGCCCCAGAAATGCAACAGTCCTGAAGGG 174
        | ||| ||| ||| ||||  |||| ||| |||| ||| ||||||| ||  || || |||
Db  149 AGCTTCCGGATCCAGTTATCAGATCATAGAAGGGTCCTCAGAAGTGTAACAGTCCTAAAGGA 208

Qy  175 CTCCCAGGCGTCGCTTCAACTGCACCGTGTCCCAGGGCTGGAGCGTCATCATGTGGGCTCT 234
        ||| ||||| | ||||||||||||||||| || |||||||| || | |||||||| | |
Db  209 CTCAGAGGCTCACTTCAACTGCACCGTGACTCACGGCTGGAAGCTTCTCATGTGGGATC 268

Qy  235 CAGTGACAATGGTGGTGCTAAGGCTGTCAGGCCCCATGGAGCCCATCATCACCAATGACCGCTT 294
        | ||| ||||||||||||||  | ||  |  | | ||  ||||||||||||||| | | ||||
Db  269 TAACCAAAATGGTGGTGCTCGAGTCTCACCACCAAGGACCCCATCATCACAACCGCTT 328

Qy  295 CACCTCTCAGAGGTACGACCCAGGGCGGGGAACTTCACCTCGGAGATGATGATCATCACAAATGT 354
        ||||| | ||| | ||  || |   | | ||   ||||||||| | ||||| |  ||||| |||
Db  329 CACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATTCATGATGT 388

Qy  355 GGAGCCCAGTGAATCGGGGAACAATCAGATGACAGCCTCCAGGAACAGTCGCCCTGCATGGATC 414
        | | |||||||| | ||||| |   | | || ||||| ||| |||| |  |  ||| | |||||
Db  389 GCAGCCCAGTGACTGCGGGATCCGTGCAATGCAGCAGAACAGCCGCCATGGGTTTGGATC 448

Qy  415 TGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTGTTCATTCCCAGTGTGTTAATCTTGTAGT 474
        |||  | ||| | ||| |||| ||||||||  | | || ||||| || | | | |  || ||||
Db  449 TGCCTTCCTTCTCAGTGCAAGTCATGGGAGACCCTGAACATTCCTAGCGAACAACTTATAGT 508

Qy  475 CGCTGAGAATGAAGACCTTGTGAAGTTACTTGTGTCTACCCTCACACTGGACCCGGCGCTT 534
        | |||||  ||||  ||| || | | || || | | || ||||||| ||||||  ||  |
Db  509 CACTGAGGGTGAACCCTGTAATGTGACTTGCATGCCCTGGCTGGACCCTCACTCCCGGA 568

Qy  535 TATTTCCTGGGAGCTCGGGTCTCCTGTGGCTAGCCATTCAAGCTATTATTTTGTTCCGGAGCC 594
        |||||||||||||| | ||| |   | | |  |||||| ||| || | | ||  ||| |||||
Db  569 TATTTCCTGGGAGCTTGAGGTTCCCGTTAAGCCATTCGAGTTAACAAACCTTATAGT 628

Qy  595 CAGGGACCTTCAAAGTGCAGTGAGCATCCTGGACCCCACAGAGCAATGGGGACTTT 654
        | ||| |||| | ||||| |||| | | | |||||| |||| | |  |||| ||||
Db  629 GGGCAACTTTATGAGGGTCTTGAGGTGTCTGGACCTCACACA-TGGCACGGACCTT 687
```

OM nucleic - nucleic search, using sw model

```
                    GenCore version 5.1.4_p5_4578
                    Copyright (c) 1993 - 2003 Compugen Ltd.
```

Run on: April 28, 2003, 23:25:02 ; Search time 2011.06 Seconds
                    (without alignments)
                    9985.963 Million cell updates/sec

Title: US-09-729-264-5
Perfect score: 1240
Sequence: 1 aggtgtgagtccagccaaca.........gtaatacaactgtagtatag 1240

Scoring table: IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0
Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                    Maximum Match 100%
                    Listing first 45 summaries

Database :

```
          EST:*
    1: em_estba:*
    2: em_esthum:*
    3: em_estin:*
    4: em_estmi:*
    5: em_estov:*
    6: em_estpl:*
    7: em_estro:*
    8: em_htc:*
    9: em_estl:*
   10: gb_est2:*
   11: gb_htc:*
   12: gb_est3:*
   13: gb_est4:*
   14: gb_est5:*
   15: em_estfun:*
   16: em_eston:*
   17: gb_gss:*
   18: em_gss_hum:*
   19: em_gss_inv:*
   20: em_gss_pln:*
   21: em_gss_vrt:*
   22: em_gss_fun:*
   23: em_gss_mam:*
   24: em_gss_mus:*
   25: em_gss_other:*
   26: em_gss_pro:*
   27: em_gss_rod:*
```

SUMMARIES

```
Result                  %    Query
No.   Score  Match  Length  DB   ID          Description
----  -----  -----  ------  --   --          -----------
 1    308.2  24.9    784    12   BG863468    BG863468 602796941
 2    278.6  22.5    794    12   BF784177    BF784177 602108039
 3    265.8  21.4    725    12   BG740428    BG740428 602633817
 4    259    20.9    631    10   BB663870    BB663870 602633817
 5    246.2  19.9    527    10   BE032610    BE032610 132035 MA
 6    197.4  15.9    517    10   BE476432    BE476432 159670 BA
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
  7    190.8  15.4    622    17   AG069679    AG069679 Pan trog1,
  8    190.8  15.4    677    17   AG107877    AG107877 Pan trog1.
  9    190    15.3    524    10   AV601192    AV601192 AV601192
 10    185.6  15.0    471     9   AA265274    AA265274 mx91d09.r
 11    185.6  15.0    423    10   AW990468    AW990468 uf37d05.y
 12    185.6  15.0    931    12   BG173684    BG173684 60233682.
 13    185.6  15.0   1552    11   AK008060    AK008060 Mus muscu
 14    184.6  14.9    486    10   AI425363    AI425363 mx91d09.y
 15    184    14.8    663    10   BE376590    BE376590 60126419.
 16    182.4  14.7    650    10   BB625217    BB625217 BB625217
 17    168.2  13.6    600    10   AZ379623    AZ379623 1M0134G22
 18    154.8  12.5    556    10   BG206666    BG206666 RST26117
 19    152.8  12.3    291    17   AI788300    AI788300 uk56d01.y
 20    151.2  12.2    530     9   AI790785    AI790785 uk28a12.y
 21    145.4  11.7    423    10   BB846577    BB846577 BB846577
 22    143.8  11.6    471    12   AZ411779    AZ411779 1M018504
 23    128.8  10.4    394    10   BF040046    BF040046 BP250004B
 24    127    10.2    471    12   BB846133    BB846133 BB846133
 25    124    10.0    685    17   AG142221    AG142221 Pan trog1
 26    110     8.9    785    13   BI452873    BI452873 603170211
 27    102.8    8.3    889    13   BI454276    BI454276 603171509
 28    100     8.1    185    14   N47851      N47851 yw95h05.r1
 29    100     8.1    234    14   N93995      N93995 za66f09.r1
 30     98.4    7.9    525    17   AQ403719    AQ403719 HS_5049_A
 31     78.6    6.3    360     9   AI647601    AI647601 uk41f12.x
 32     78.6    6.3    374     9   AI648720    AI648720 uk28a12.x
 33     78.6    6.3    406     9   AI788973    AI788973 uk56d01.x
 34     77      6.2    644    10   BB531388    BB531388 BB531388
 35     67.8    5.5    397    10   AW112084    AW112084 MC7851 mo
 36     63.2    5.1    282    10   BB564363    BB564363 BB564363
 37     58.4    4.7    466    17   AQ224346    AQ224346 HS_2258_B
 38     56      4.5    818    12   BG445021    BG445021 GA__Ea002
 39     52.6    4.2    422     9   AI756062    AI756062 E1ESTea09
 40     52.4    4.2    458    10   AL514929    AL514929 AL514929
 41     50.6    4.1    639    17   AG113259    AG113259 Pan trog1
 42     50      4.0    427    17   AO595624    AO595624 HS_5440_B
 43     50      4.0    447    10   AV431233    AV431233 AV431233
 44     49.8    4.0    428     9   AA695064    AA695064 GM02059.5
 45     49.6    4.0    925    17   CNS03381E   CNS03381E Tetraodon
```

ALIGNMENTS

RESULT 1
BG863468
LOCUS   BG863468   784 bp   mRNA   linear
DEFINITION 602796941F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4917803 5',
            mRNA sequence.
ACCESSION   BG863468
VERSION   BG863468.1   GI:14214006
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 784)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/;
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10830 row: b column: 12
            High quality sequence stop: 768.
FEATURES   Location/Qualifiers

```
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/185,432
;   FILING DATE: 21-JAN-1994
;   CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
;   NAME: Misrock, S. Leslie
;   REGISTRATION NUMBER: 18,872
;   REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 790-9090
;   TELEFAX: (212) 869-8864/9741
;   TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 5063 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: unknown
;   MOLECULE TYPE: cDNA/DNA (genomic)
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 864..1349
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1944..3370
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 3439..3736
US-08-185-432-1
```

```
Query Match        3.5%;   Score 41;  DB 1;  Length 5063;
Best Local Similarity   66.3%;  Pred. No. 0.076;
Matches  59;  Conservative  0;  Mismatches  30;  Indels  0;  Gaps  0;

Qy   801  GCTGCTGCTGCTGCCGCGCCGGTCGTGCTTGTTGTGGCTTGCAAACTGCTGCC  860
          ||||||||||||||||| || ||  | ||||  |||| |||||||||| || ||
Db  2988  GCTGCTGCTGCCGCCTTGTTGCTCGTCTGCTTGTTGTTGTTGTTGCTGCT  2929

Qy   861  GCTGTGTAGAAGAAAAAGAGAGAGAGAGAGAGATTTCGTGTATT  889
          ||||| |||| | |||  ||
Db  2928  GCTGTTGCTTGTCTGGGCGATGCCAGTGGGTAAT  2900
```

```
RESULT 15
US-08-479-913E-1/c
Sequence 1, Application US/08479913E
; Patent No. 6416998
; GENERAL INFORMATION:
;   APPLICANT: O'Malley, Bert W.
;   APPLICANT: Tsai, Ming-Jer
;   APPLICANT: Ledebur, Harry C.
;   APPLICANT: Kittle, Joseph D.
;   TITLE OF INVENTION: MODIFIED STEROID HORMONES FOR GENE THERAPY AND METHODS FOR THEIR
;   FILE REFERENCE: 212/133
;   CURRENT APPLICATION NUMBER: US/08/479,913E
;   CURRENT FILING DATE: 1995-06-07
;   PRIOR APPLICATION NUMBER: US 07/939,246
;   PRIOR FILING DATE: 1992-09-02
;   NUMBER OF SEQ ID NOS: 6
;   SOFTWARE: Microsoft Word 97
; SEQ ID NO 1
;   LENGTH: 6177
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;     NAME/KEY: misc_feature
;     OTHER INFORMATION: Plasmid pGRO403R
US-08-479-913E-1
```

```
Query Match        3.5%;   Score 40.8;  DB 4;  Length 6177;
Best Local Similarity   75.0%;  Pred. No. 0.097;
Matches  51;  Conservative  0;  Mismatches  17;  Indels  0;  Gaps  0;
```

```
Qy   801  GCTGCTGCTGCTGCCGCGCCGGTCGTGCTTGTTGTGGCTGCAAACTGCTGCCGTTGTTGTTTCT  860
          |||||||||||| |||| ||||| |||||| | |||| ||||| ||||||||||| |||  ||||| ||
Db  4785  GCTGCCTGCTGCTGCTCGTCGTGCTGCTGGCTGCCTGCCGTTGTTGTTGTTTGCGCT  4726

Qy   861  GCTGTAGA  868
          ||| | ||
Db  4725  GCTGCACA  4718
```

```
Search completed: April 29, 2003, 10:49:49
Job time : 81.8976 secs
```

RESULT 12
US-08-185-432-3/c
; Sequence 3, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
;   APPLICANT: Artavanis-Tsakonas, Spyridon
;   APPLICANT: Busseau, Isabelle
;   APPLICANT: Diederich, Robert J.
;   APPLICANT: Xu, Tian
;   APPLICANT: Matsuno, Kenji
;   TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND RELATED METHODS AND COMPOSITIONS
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PENNIE & EDMONDS
;   STREET: 1155 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10036-2711
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/185,432
;   FILING DATE: 21-JAN-1994
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Misrock, S. Leslie
;   REGISTRATION NUMBER: 18,872
;   REFERENCE/DOCKET NUMBER: 7326-006
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 790-9090
;   TELEFAX: (212) 869-8864/9741
;   TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3771 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: unknown
;   MOLECULE TYPE: cDNA
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 345..2558
;
US-08-185-432-3

Query Match   3.5%;   Score 41;   DB 1;   Length 3771;
Best Local Similarity   66.3%;   Pred. No. 0.064;
Matches   59;   Conservative   0;   Mismatches   30;   Indels   0;   Gaps   0;

Qy   801 GCTGCTGCTGCTGCCGCCGTGCGTTGTTGTTGGGCTGCACAGCTGCCGGCGTCGCCGTTG TTTCT 860
      |||||||||| |||| |||||||| ||||||||||||| |||||||||| |||||||| ||
Db 1875 GCTGCTGCTGCGGCGCCTGTTGCCTGTTGCCATCTGCAGCGTGGTTGCCATCTGAC GTTGGCTGCT 1816

Qy   861 GCTGTAGGAAGAAAAAGAGGAGATTTCGTATT 889
      |||||| | ||| ||||||| || |||||
Db 1815 GCTGTTGCTGCTGCCGGGATGCAGTGGTAAT 1787

RESULT 13
US-08-185-432-23
; Sequence 23, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
;   APPLICANT: Artavanis-Tsakonas, Spyridon
;   APPLICANT: Busseau, Isabelle
;   APPLICANT: Diederich, Robert J.
;   APPLICANT: Xu, Tian
;   APPLICANT: Matsuno, Kenji
;   TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND

;   TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PENNIE & EDMONDS
;   STREET: 1155 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10036-2711
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/185,432
;   FILING DATE: 21-JAN-1994
;   CLASSIFICATION: 530
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Misrock, S. Leslie
;   REGISTRATION NUMBER: 18,872
;   REFERENCE/DOCKET NUMBER: 7326-006
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (212) 790-9090
;   TELEFAX: (212) 869-8864/9741
;   TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 23:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 3771 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: unknown
;   MOLECULE TYPE: cDNA
;
US-08-185-432-23

Query Match   3.5%;   Score 41;   DB 1;   Length 3771;
Best Local Similarity   66.3%;   Pred. No. 0.064;
Matches   59;   Conservative   0;   Mismatches   30;   Indels   0;   Gaps   0;

Qy   801 GCTGCTGCTGCTGCCGCCGTGCGTTGTTGTTGGGCTGCACAGCTGCCGGCGTCGCCGTTG TTTCT 860
      |||||||||| |||| |||||||| ||||||||||||| |||||||||| |||||||| ||
Db 1897 GCTGCTGCTGCGGCGCCTGTTGCCTGTTGCCATCTGCAGCGTGGTTGCCATCTGAC GTTGGCTGCT 1956

Qy   861 GCTGTAGGAAGAAAAAGAGGAGATTTCGTATT 889
      |||||| | ||| ||||||| || |||||
Db 1957 GCTGTTGCTGCTGCCGGGATGCAGTGGTAAT 1985

RESULT 14
US-08-185-432-1/c
; Sequence 1, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
;   APPLICANT: Artavanis-Tsakonas, Spyridon
;   APPLICANT: Busseau, Isabelle
;   APPLICANT: Diederich, Robert J.
;   APPLICANT: Xu, Tian
;   APPLICANT: Matsuno, Kenji
;   TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND RELATED METHODS AND COMPOSITIONS
;   NUMBER OF SEQUENCES: 23
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: PENNIE & EDMONDS
;   STREET: 1155 Avenue of the Americas
;   CITY: New York
;   STATE: New York
;   COUNTRY: U.S.A.
;   ZIP: 10036-2711
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30

PCT-US92-11107-12

Query Match
Best Local Similarity   64.9%;   Score 42.6;   DB 5;   Length 3292;
Matches 63;   Conservative   0;   Mismatches 34;   Indels   0;   Gaps   0;

```
Qy    797  ATACGCTGCTGCTCGTCGTGCCGCGTCGCCAACTGTCGTGCTCCCGTTGTTGT  856
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    818  AGATGATGATGTGTTGTCTTGAAGTTGTTGTTGTTGCAACTGGTGTGTTGCTGCTGTAGC  759

Qy    857  TTCTGCTGTAGAAGAAAAAGAGGATTTCGTATTCAAT  893
            |||||||||||||||||||||||||||||||||||
Db    758  TGCTGCTGTTGCTCTGCAACTGCTGTTGTTGTTAATT  722
```

RESULT 9
US-09-043-303-12/c
; Sequence 12, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
;   APPLICANT: TSUJI, Shoji
;   APPLICANT: SANPEI, Kazujiro
;   TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
;   TITLE OF INVENTION: Primers Therefor
;   FILE REFERENCE: 0760-0241P
;   CURRENT APPLICATION NUMBER: US/09/043,303
;   CURRENT FILING DATE: 1998-05-18
;   EARLIER APPLICATION NUMBER: PCT/JP96/01999
;   EARLIER FILING DATE: 1996-07-18
;   NUMBER OF SEQ ID NOS: 17
;   SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 78
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-043-303-12

Query Match
Best Local Similarity   78.1%;   Score 41.6;   DB 4;   Length 78;
Matches 50;   Conservative   0;   Mismatches 14;   Indels   0;   Gaps   0;

```
Qy    801  GCTGCTGCTGCTGCTGCCGCGTCGTTGTTGTTGGCGCTGCAACTGCTGCTCCCGTTGTTGT  860
            ||||||||||| |||||||||||||||||||||| ||||||||||||| |||||||
Db     67  GCTGCTGCTGCTCGTCGTGCTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  8

Qy    861  GCTG 864
            ||||
Db      7  GCTG 4
```

RESULT 10
US-09-043-303-7/c
; Sequence 7, Application US/09043303
; Patent No. 6251589
; GENERAL INFORMATION:
;   APPLICANT: TSUJI, Shoji
;   APPLICANT: SANPEI, Kazujiro
;   TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
;   TITLE OF INVENTION: Primers Therefor
;   FILE REFERENCE: 0760-0241P
;   CURRENT APPLICATION NUMBER: US/09/043,303
;   CURRENT FILING DATE: 1998-05-18
;   EARLIER APPLICATION NUMBER: PCT/JP96/01999
;   EARLIER FILING DATE: 1996-07-18
;   NUMBER OF SEQ ID NOS: 17
;   SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 203
;   TYPE: DNA
;   ORGANISM: p-2093 plasmid
US-09-043-303-7

Query Match   3.68;   Score 41.6;   DB 4;   Length 203;

Best Local Similarity   78.1%;   Pred. No. 0.0081;
Matches 50;   Conservative   0;   Mismatches 14;   Indels   0;   Gaps   0;

```
Qy    801  GCTGCTGCTGCTGCTGCCGCGTCGTCGTTGTTGGCTGGTTGTTGGCTGCAACTGCTGTTGTTGTTCT  860
            ||||||||||| |||||||||||||||| || |||||||||||| || ||||| |||| ||
Db     67  GCTGCTGCTGCTCGTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  8

Qy    861  GCTG 864
            ||||
Db      7  GGTG 4
```

RESULT 11
US-08-145-617-5/c
; Patent No. 5766847
; Sequence 5, Application US/08145617
; GENERAL INFORMATION:
;   APPLICANT: Jackle, Herbert
;   APPLICANT: Tautz, Diethard
;   TITLE OF INVENTION: PROCESS FOR ANALYZING LENGTH
;   TITLE OF INVENTION: POLYMORPHISMS IN DNA REGIONS
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
;   STREET: 301 N. Washington Street, P.O. Box 747
;   CITY: Falls Church
;   STATE: Virginia
;   COUNTRY: United States of America
;   ZIP: 22046
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/145,617
;   FILING DATE:
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/681,494
;   FILING DATE: 10-JUN-1991
;   APPLICATION NUMBER: DE P3834636.2
;   FILING DATE: 11-OCT-1988
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Svensson, Leonard R.
;   REGISTRATION NUMBER: 30,330
;   REFERENCE/DOCKET NUMBER: 147-122PCT
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 703-241-1300
;   TELEFAX: 703-241-2848
;   TELEX: 248345
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 379 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-08-145-617-5

Query Match
Best Local Similarity   67.4%;   Score 41.2;   DB 1;   Length 379;
Matches 58;   Conservative   0;   Mismatches 28;   Indels   0;   Gaps   0;

```
Qy    779  CCGACGTGTACTCTTACAATACGCTGCTGCCGCCGTCGTTGTTGTTGGGCTGCAAC  838
            |||| |||| |||||||||||||||||||||||||||||| ||||||| ||||
Db    214  CCGAACTCCAGGCCTCCGAGTTGCTGCTGCTGCTGTTGCTGCTGCTGCTGCTGCT  155

Qy    839  TGCTGCTGCCGTTGTTGTTTCTGCTG  864
            |||||| ||| ||| |||| ||||
Db    154  TGCTGCTGCTGATGTTGCTGCTGCTGTTG  129
```

```
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..1626
US-08-258-442-12

Qy 797 ATACGCGTCGCTGCGTCCCGCGGTCGTTGTTGTGGGGTCGCCAACTGCGTCCGGTTGTTGT 856
       | ||  | || || | ||||||||  | || ||  |||||| || ||| || |||||||
Db 818 AGATGATGATGTTGTTGTTGAAGTTGTTGTTGTTGTTGCAACTGGTGTTGCCTGTTGTAGC 759

Qy 857 TTCTGCTGTAGGAAGAAAAGAGGATTTCGTATTCAAT 893
       ||||| |||  |||| | |||| || ||||| ||| |
Db 758 TGCTGCGTTGCGTGCGCAACTGCGTGTTGTTGTGTAATT 722

Query Match        64.9%;  Score 42.6;  DB 1;  Length 3292;
Best Local Similarity  64.9%;  Pred. No. 0.02;
Matches  63;  Conservative   0;  Mismatches  34;  Indels   0;  Gaps   0;
```

```
RESULT 7
US-08-328-809-7/c
Sequence 7, Application US/08328809
Patent No. 5705334
GENERAL INFORMATION:
   APPLICANT: Lippard, Stephen J.
   APPLICANT: Essigmann, John M.
   APPLICANT: Donahue, Brian A.
   APPLICANT: Toney, Jeffrey H.
   APPLICANT: Bruhn, Suzanne L.
   APPLICANT: Pil, Pieter M.
   APPLICANT: Brown, Steven
   APPLICANT: Kellett, Patti
   TITLE OF INVENTION: Uses For DNA Structure-Specific
   TITLE OF INVENTION: Recognition Proteins
   NUMBER OF SEQUENCES: 8
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: Patent Administrator, Testa, Hurwitz & Thibeault
      STREET: 53 State Street
      CITY: Boston
      STATE: MA
      COUNTRY: USA
      ZIP: 02109
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/328,809
      FILING DATE:
   CLASSIFICATION: 435
   ATTORNEY/AGENT INFORMATION:
      NAME: Fenton, Gillian M.
      REGISTRATION NUMBER: 36,508
      REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: 617-248-7000
      TELEFAX: 617-248-7100
INFORMATION FOR SEQ ID NO: 7:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 3292 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
   MOLECULE TYPE: DNA (genomic)
   ORIGINAL SOURCE:
      ORGANISM: Saccharomyces cerevisiae
   IMMEDIATE SOURCE:
      CLONE: lambda ypt
   MOLECULE TYPE: DNA (genomic)
   ORIGINAL SOURCE:
      ORGANISM: Saccharomyces cerevisiae
   IMMEDIATE SOURCE:
      CLONE: lambda ypt
   FEATURE:
      NAME/KEY: CDS
      LOCATION: 1..1626
```

```
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 1..1626
US-08-328-809-7

Qy 797 ATACGCGTCGCTGCGTCCCGCGGTCGTTGTTGTGGGGTCGCCAACTGCGTCCGGTTGTTGT 856
       | ||  | || || | ||||||||  | || ||  |||||| || ||| || |||||||
Db 818 AGATGATGATGTTGTTGTTGAAGTTGTTGTTGTTGTTGCAACTGGTGTTGCCTGTTGTAGC 759

Qy 857 TTCTGCTGTAGGAAGAAAAGAGGATTTCGTATTCAAT 893
       ||||| |||  |||| | |||| || ||||| ||| |
Db 758 TGCTGCGTTGCGTGCGCAACTGCGTGTTGTTGTGTAATT 722

Query Match        64.9%;  Score 42.6;  DB 1;  Length 3292;
Best Local Similarity  64.9%;  Pred. No. 0.02;
Matches  63;  Conservative   0;  Mismatches  34;  Indels   0;  Gaps   0;
```

```
RESULT 8
PCT-US92-11107-12/c
Sequence 12, Application PCT/US9211107
GENERAL INFORMATION:
   APPLICANT: Donahue, Brian A.
   APPLICANT: Toney, Jeffrey H.
   APPLICANT: Bruhn, Suzanne L.
   APPLICANT: Pil, Pieter M.
   APPLICANT: Brown, Steven
   APPLICANT: Kellett, Patti
   APPLICANT: Essigmann, John M.
   APPLICANT: Lippard, Stephen J.
   TITLE OF INVENTION: DNA Structure Specific Recognition
   TITLE OF INVENTION: Protein and Uses Therefor
   NUMBER OF SEQUENCES: 13
   CORRESPONDENCE ADDRESS:
      ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
      STREET: 2 Militia Drive
      CITY: Lexington
      STATE: MA
      COUNTRY: USA
      ZIP: 02173
   COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
      APPLICATION NUMBER: PCT/US92/11107
      FILING DATE: 19921218
   PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 07/539,906
      FILING DATE: 18-JUN-1990
   ATTORNEY/AGENT INFORMATION:
      NAME: Granahan, Patricia
      REGISTRATION NUMBER: 32,227
      REFERENCE/DOCKET NUMBER: MIT-4787AAA
   TELECOMMUNICATION INFORMATION:
      TELEPHONE: 617-861-6240
      TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 12:
   SEQUENCE CHARACTERISTICS:
      LENGTH: 3292 base pairs
      TYPE: NUCLEIC ACID
      STRANDEDNESS: double
      TOPOLOGY: linear
   MOLECULE TYPE: DNA (genomic)
   ORIGINAL SOURCE:
      ORGANISM: Saccharomyces cerevisiae
   IMMEDIATE SOURCE:
      CLONE: lambda ypt
   FEATURE:
      NAME/KEY: CDS
      LOCATION: 1..1626
```

RESULT 5
US-07-814-964-12/c
; Sequence 12, Application US/07814964
; Patent No. 5359047
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellett, Patti
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/814,964
; FILING DATE: 19911226
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3292 base pairs
; TYPE: NUCLEIC ACID

Query Match   3.7%; Score 42.8; DB 4; Length 1776;
Best Local Similarity   61.8%; Pred. No. 0.012;
Matches   68; Conservative   0; Mismatches   42; Indels   0; Gaps   0;

Qy   801 GCTGGCTGCTGCTGCGCGCCGTCGTTGTTGTGGCTGCCAACTGCGTGCGCCTTGTTGTTCT 860
          ||||||||||||||||| ||||||||||| |||||||||| ||||||||||||||||||||
Db   975 GCTGGCTGCTGCTGCTGCTGCTGCTGCTGCGCTGCTGCTGCTGGTTGCCAACTGGTGTTGC 916

Qy   861 GCTGTAGAAGAAAAGAGGGATTTCGTATTCAATTTCAAAAGAAATCTGAA 910
          |||| | || || || || |||||||
Db   915 GCTGCTGTTTTTCAAAGTAGGCTTCTCGTCTCTCCGAAGCTCTTCTGAA 866

US-09-041-886-12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1116
; MOLECULE TYPE: DNA (genomic)
; TOPOLOGY: linear
; STRANDEDNESS: single
; TYPE: nucleic acid
; LENGTH: 1776 base pairs
; SEQUENCE CHARACTERISTICS:
; INFORMATION FOR SEQ ID NO: 12:
; TELEFAX: (619) 535-8949
; TELEPHONE: (619) 535-9001

RESULT 6
US-08-258-442-12/c
; Sequence 12, Application US/08258442
; Patent No. 5670621
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellett, Patti
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,442
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3292 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

Query Match   3.6%; Score 42.6; DB 1; Length 3292;
Best Local Similarity   64.9%; Pred. No. 0.02;
Matches   63; Conservative   0; Mismatches   34; Indels   0; Gaps   0;

Qy   797 ATACGCTGCGTGCCTGCGCGCGCCGCCGTTGTTGGTGGCTGCAACTGCGTGCCGGTTGTTGT 856
          | |||||||| |||||||||| ||||| ||||||||| ||||||||||| |||||| |||||
Db   818 AGATGATGATGTTGTTGTTGTGAAGTTGTTGTTGTTGTTGCAACTGGTGTTGTTGCTGTTGAGC 759

Qy   857 TTCTGCTGTAGAAGAAAAGAGGAGATTTCGTATTCAAT 893
          | |||||| | || || || ||
Db   758 TGCTGCTGTTGCTGCGAACTGCTGTTGTTGTTGTAATT 722

US-07-814-964-12
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1626
; CLONE: lambda ypt
; IMMEDIATE SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; ORIGINAL SOURCE:
; MOLECULE TYPE: DNA (genomic)
; TOPOLOGY: linear
; STRANDEDNESS: double

STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/531,927B
FILING DATE: 21-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP H6-251600
FILING DATE: 21-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: ATH95-01A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 325 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: modified_base
LOCATION: 145
FEATURE:
NAME/KEY: modified_base
LOCATION: 194

Query Match   4.1%; Score 48.2; DB 2; Length 325;
Best Local Similarity 57.8%; Pred. No. 0.00012;
Matches 100; Conservative 2; Mismatches 70; Indels 1; Gaps 1;

```
Qy  785 TGTACTCTTACAATACGCTGCTGCTGCCGTCGTGTTGTGCTGCAACTGCTGC 844
            ||| || | |||| ||||||||||||||||||||||||||| ||||| |||
Db  210 TGTCCTGATAGGTCCCSCTGCTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGC 151

Qy  845 TGCCGTTGTTGTTCTGCCTGGTAGAA-GAAAAAGAGGA 903
            ||| || || ||||| |  |||||  ||||||||||
Db  150 TGCTGGYTGCCTGCTTTTGCTGCTGTCTGAAACATTCA 91

Qy  904 ATCTGAAAAAGAGAAGACAAAAAGAAAAGAAAAACTGAGACAGAAAAGTGAAATGAA 956
            ||||||| |||| ||| |  |||||| |||||||||||| |||| ||||||||||
Db   90 AAACTTAAAAGAATAAAATACACCATGAGAAAAAACTATTCATAAGGAAAAATACA 38
```

RESULT 3
US-08-531-927B-1/c
Sequence 1, Application US/08531927B
Patent No. 5840491
GENERAL INFORMATION:
APPLICANT: Kakizuka, Akira
TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
Patent No. 5840491
TITLE OF INVENTION: Disease Gene and Uses Thereof
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/531,927B
FILING DATE: 21-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP H6-251600
FILING DATE: 21-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: ATH95-01A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1776 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 36..1115

Query Match   3.7%; Score 42.8; DB 2; Length 1776;
Best Local Similarity 61.8%; Pred. No. 0.012;
Matches 68; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```
Qy  801 GCTGCYGCTGCCYGCCGCGCGTCGTTGTTGGGCTGCAACTGCTGCCGCCGTTGTTGTTTCT 860
            |||| ||||| ||| | || | ||||||  |||||||||||| | |||| || || |||
Db  975 GCTGCTCCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGTTGTTGGCTGCTGCTGCTTTT 916

Qy  861 GCTGTAGAAGAAAAAGAGGATTTCGTATTCAATTTCAAAAGAAAATCTGAA 910
            |||| | | ||| | || ||| |||  |||| | |  |||  | ||||||
Db  915 GCTGCTGCTTTTTCAAAGTAGGCTTCTCGTTCTTCCGAAGCTCTTCTGAA 866
```

RESULT 4
US-09-041-886-12/c
Sequence 12, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
APPLICANT: Bredesen, Dale E.
APPLICANT: Rabizadeh, Shahroz
TITLE OF INVENTION: Proapoptotic Peptides, Dependence
TITLE OF INVENTION: Polypeptides and Methods of Use
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: United States
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,886
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LJ 2626
TELECOMMUNICATION INFORMATION:

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2003, 23:25:03 ; Search time 48.8976 Seconds
                       7325.486 Million cell updates/sec

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
               1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
               2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
               3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
               4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
               5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
               6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Searched: 441362 seqs, 15338381 residues

Title: US-09-729-264-3
Perfect score: 1168
Sequence: 1 agtgatcatggtggcaggag...............gtaatacaactgtgagtatag 1168

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	51.8	4.4	397	3	US-09-253-691-3	Sequence 3, Appli
2	48.2	4.1	325	2	US-08-531-927B-3	Sequence 3, Appli
3	42.8	3.7	1776	4	US-08-531-927B-1	Sequence 1, Appli
4	42.8	3.7	1776	4	US-09-041-886-12	Sequence 12, Appl
5	42.6	3.6	3292	1	US-07-814-964-12	Sequence 12, Appl
6	42.6	3.6	3292	1	US-08-258-442-12	Sequence 12, Appl
7	42.6	3.6	3292	1	US-08-328-809-7	Sequence 7, Appl
8	42.6	3.6	3292	5	PCT-US92-11107-12	Sequence 12, Appl
9	41.6	3.6	78	4	US-09-043-303-12	Sequence 12, Appl
10	41.6	3.6	203	4	US-09-043-303-7	Sequence 7, Appl
11	41.2	3.6	379	1	US-08-145-617-5	Sequence 5, Appli
12	41	3.5	3771	1	US-08-185-432-3	Sequence 3, Appli
13	41	3.5	3771	1	US-08-185-432-23	Sequence 23, Appl
14	41	3.5	5063	1	US-09-185-432-1	Sequence 1, Appli
15	40.8	3.5	6177	4	US-08-479-913E-1	Sequence 1, Appli
16	40.8	3.5	7257	4	US-09-091-042A-1	Sequence 1, Appli
17	40.4	3.5	484	6	5496550-5	Patent No. 5496550
18	40.4	3.5	2674	4	US-09-817-180-1	Sequence 1, Appli
19	40.4	3.5	10348	2	US-08-457-273B-41	Sequence 41, Appl
20	40.4	3.5	10348	3	US-08-556-419-13	Sequence 13, Appl
21	40.4	3.5	10348	1	US-09-041-886-14	Sequence 14, Appl
22	40.4	3.5	10366	1	US-08-246-982A-5	Sequence 5, Appli
23	40.4	3.5	10366	1	US-08-453-265-5	Sequence 5, Appli
24	40	3.5	75	4	US-09-043-303-10	Sequence 10, Appl
25	40	3.4	78	4	US-09-043-303-11	Sequence 11, Appl
26	40	3.4	154	1	US-08-469-802B-6	Sequence 6, Appli
27	40	3.4	154	2	US-08-267-803B-6	Sequence 6, Appli
28	40	3.4	165	4	US-09-043-303-17	Sequence 17, Appl
29	40	3.4	168	1	US-08-469-802B-4	Sequence 4, Appli
30	40	3.4	168	2	US-08-267-803B-4	Sequence 4, Appli
31	40	3.4	171	1	US-08-469-802B-5	Sequence 5, Appli
32	40	3.4	171	2	US-08-267-803B-5	Sequence 5, Appli
33	40	3.4	195	1	US-08-469-802B-2	Sequence 2, Appli
34	40	3.4	195	2	US-08-267-803B-2	Sequence 2, Appli
35	40	3.4	234	1	US-08-469-802B-3	Sequence 3, Appli
36	40	3.4	234	2	US-08-267-803B-3	Sequence 3, Appli
37	40	3.4	270	1	US-09-146-054-8	Sequence 8, Appli
38	40	3.4	270	4	US-09-664-977A-8	Sequence 8, Appli
39	40	3.4	477	1	US-09-135-994-1	Sequence 1, Appli
40	40	3.4	506	2	US-08-469-802B-7	Sequence 7, Appli
41	40	3.4	506	1	US-08-267-803B-7	Sequence 7, Appli
42	40	3.4	623	4	US-09-043-303-5	Sequence 5, Appli
43	40	3.4	3366	1	US-08-469-802B-1	Sequence 1, Appli
44	40	3.4	3366	2	US-08-267-803B-1	Sequence 1, Appli
45	40	3.4	3376	1	US-08-320-559-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-09-253-691-3/c
Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; TITLE OF INVENTION: Using Trinucleotide Repeats Sequence
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; CURRENT FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
US-09-253-691-3

  Query Match   4.4%; Score 51.8; DB 3; Length 397;
  Best Local Similarity   58.9%; Pred. No. 1.2e-05;
  Matches   89; Conservative   0; Mismatches 62; Indels   0; Gaps   0;

Qy   801 GCTGCTGCCTGCTGCCGCGCCGTGCGTTGTTGTGTGGCGCAACTGCGTGCTGCCGTTGTTTCT 860
              |||||||||||||||| || | ||||| || |||||||||||||||||||| || ||||
Db   200 GCTGCTCGCTGCGCTGCTGCTGCCGTTGCTGCTCTGCTCGCTGCCGTTGCTGCTCGCTGCT 141

Qy   861 GCTGTAGAAGARAAAAGAGGGATTTCGTATTCAATTTCAAAAGAAATCTGAAAAAGAGAAGA 920
              ||||| |||| | || || | | | || |||| ||| | | | | |||| || ||
Db   140 GCTGTTGCCTGCGTTTTGCTGCTCGTCGAAACATTCAAAAGTGAAGTATATTTAAAAAACAA 81

Qy   921 CAAACAAAGAAAACTGAGACAGAAAGTGGAAA 951
              || | | | | | || | | | ||||
Db   80 AACTTAAAAGAATAAAATACACCATGAGAAAA 50

RESULT 2
US-08-531-927B-3/c
Sequence 3, Application US/08531927B
; Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491
; TITLE OF INVENTION: Disease Gene and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

DR　WPI; 2000-662424/64.
XX
PT　Plasmid vector arraying neuropsychiatric gene more than copy number
PT　containing trinucleotide repeats its use thereof -
XX
PS　Disclosure; Page 12; 23pp; Korean.
XX
CC　This invention relates to the use of a plasmid vector array
CC　containing a neuropsychiatric gene containing trinucleotide repeats.
CC　Tri-nucleotide repeats and their expansion areknown to be involved
CC　in a number of diseases including fragile X syndrome, spinocerebellar
CC　ataxia type III, hereditary hypertrophic, cardiomyopathy, dentatorubral
CC　and pallidoluysian atrophy (DRPLA), spinocerebellar ataxia, Marfan
CC　syndrome; X-linked spinobubar atrophy. The present sequence
CC　represents a nucleotide sequence of the invention showing the
CC　the CAG tri-nucleotide repeats.
XX
SQ　Sequence 403 BP; 121 A; 108 C; 105 G; 69 T; 0 other;

Query Match　　4.4%;　Score 51.8;　DB 21;　Length 403;
Best Local Similarity　58.9%;　Pred. No. 0.00036;
Matches　89;　Conservative　0;　Mismatches　62;　Indels　0;　Gaps　0;

Qy　801 GCTGCTGCTGCTGCTGCCGCGTCGTTGTTGTTGGGCTGCAACTGCTGCTGCCGTTGTTGTTTCT 860
         ||||||||||||||||||||||| |||||| ||||||||||||||| ||||||| |||| |
Db　215 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156

Qy　861 GCTGTAGAAGAAGAAAAAGGATTTCGTATTCAAATTTCAAAAGAAAATCTGAAAAAAGAGAAGA 920
         ||||||| |||| ||||||| || |||||||||||||||||||||| |||||| |||||||
Db　155 GCTGTTGCTGCTGCTTTTGCTGCTCGTCTGAAACATTCAAAAGTGAAGTATATTTAAAAACAA 96

Qy　921 CAAACAAAGAAAACTGAGACAGAGAAAGTGGAAA 951
         ||| ||| || ||| |||| |||| ||||
Db　95 AACTTAAAAGAATAAAATACACCATGAGAAAA 65

XX
CC　capable of detecting 1000 or more genes from Drosophila. The invention is
CC　useful in developmental biology and in elucidating cell signalling and
CC　cell-cell interactions in higher eukaryotes for the development of
CC　insecticides, therapeutics and pharmaceutical drugs. The invention
CC　discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC　sequences (ABL01840-ABL16175) and the encoded proteins
CC　(ABB57737-ABB72072).
CC　The sequence data for this patent did not form part of the printed
CC　specification, but was obtained in electronic format directly from WIPO
CC　at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ　Sequence 46954 BP; 13370 A; 9970 C; 9854 G; 13760 T; 0 other;

Query Match　　4.2%;　Score 49.4;　DB 23;　Length 46954;
Best Local Similarity　63.6%;　Pred. No. 0.021;
Matches　91;　Conservative　0;　Mismatches　51;　Indels　1;　Gaps　1;

Qy　764 ACCATGCTTCTTGACGCCGACGTGTACTCT-TACAATACGCTGCTCGTCGCGGCCGCCGTCG 822
         || ||| ||||| || ||| |||| |||| ||||||| ||| |||||| ||||||| |||
Db　16836 ACGATGGTGCTGCCGATGATGTGGATGCCTCTGGATGCCTGGCTGGCTGCTGCCAGATGCAA 16895

Qy　823 TTGTTTGTGGCTGCAACTGCTGCTGCCGTTGTTGTTTCTGCTGTAGAAGAGAAAAAGAGGATT 882
         ||||| |||||| |||||| ||||||||| ||||||| || ||| ||||||| ||| || |||
Db　16896 TTGCTGCCGGCTGCTGCCCTGCTGCTGCTGCTGCTGCTGCTGCCAGTTGCGGATGCGGATGGGGATGCGGATGAGTGCA 16955

Qy　883 TCGTATTCAATTTCAAAAGAAAT 905
         | ||||||||||||||||||||
Db　16956 AAAGTTACAAATTTCAAAACAAAT 16978

Search completed: April 29, 2003, 11:05:24
Job time : 342.628 secs

RESULT 15
ABL16830
ID　ABL16830 standard; DNA; 46954 BP.
XX
AC　ABL16830;
XX
DT　26-MAR-2002　(first entry)
XX
DE　Drosophila melanogaster genomic polynucleotide SEQ ID NO 1963.
XX
KW　Drosophila; developmental biology; cell signalling; insecticide;
KW　pharmaceutical; gene; ds.
XX
OS　Drosophila melanogaster.
XX
PN　WO200171042-A2.
XX
PD　27-SEP-2001.
XX
PF　23-MAR-2001; 2001WO-US09231.
XX
PR　23-MAR-2000; 2000US-191637P.
PR　11-JUL-2000; 2000US-0614150.
XX
PA　(PEKE) PE CORP NY.
XX
PI　Venter JC, Adams M, Li PWD, Myers EW;
XX
DR　WPI; 2001-656860/75.
XX
PT　New isolated nucleic acid detection reagent for detecting 1000 or more
PT　genes from Drosophila and for elucidating cell signalling and cell-cell
PT　interactions -
XX
PS　Claim 1; SEQ ID NO 1963; 21pp + Sequence Listing; English.
XX
CC　The invention relates to an isolated nucleic acid detection reagent

XX
CC   This invention relates to the use of a plasmid vector array
CC   containing a neuropsychiatric gene containing trinucleotide repeats.
CC   Tri-nucleotide repeats and their expansion are known to be involved
CC   in a number of diseases including fragile X syndrome, spinocerebellar
CC   ataxia type III, hereditary hypertrophic, cardiomyopathy, dentatorubral
CC   and pallidoluysian atrophy (DRPLA), spinocerebellar ataxia, Marfan
CC   syndrome; X-linked spinobubar atrophy. The present sequence represents
CC   a tri-nucleotide repeat related sequence #5 disclosed in the scope
CC   of the invention.
CC   Note: Nucleotides which were illegible in the specification have
CC   been represented by an n to keep nucleotide numbering correct.
XX
SQ   Sequence 389 BP; 107 A; 92 C; 91 G; 93 T; 6 other;

Query Match   Best Local Similarity   57.4%;
  Score 51.8; DB 21; Length 389;
  Matches   89; Conservative   0; Mismatches   66; Indels   0; Gaps   0;

Qy   801 GCTGCTGCTGCTGCCGCGCCGTCGTTGTTGTTGGCTGCAACTGCGTTGTTGTTGTTTCT 860
           ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   178 GCTGCTGCTGCTGCTGTTGCTGCTGTTGCTGCTGCTGCTGCTGTTTGTTGTTGTTGCT   119

Qy   861 GCTGTAGAGAAGAAAAAGAGGGATTTCGTATTCAATTTCAAATTTGAAAAAGAGAGAGA 920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   118 GCTGTTGCTGCTTTTGCTGCTGTCTGAAAACATTCAAAAAGTGAAGTATATTTAAAAAACAA   59

Qy   921 CAAACAAAGAAAACTGAGACAGAAGAAAGTGGAAAATGAA   955
           |||||||||||||||||||||||||||||||||||
Db   58 AANTTAAAAGAAGAANACCTACACCATGAGNTANANAA   24

RESULT 13
AAX89891/c
ID   AAX89891 standard; DNA; 397 BP.
XX
AC   AAX89891;
XX
DT   05-NOV-1999   (first entry)
XX
DE   Spinocerebellar ataxia type III (SCAIII) gene fragment.
XX
KW   Spinocerebellar ataxia type III; SCAIII; reverse dot hybridisation;
KW   PCR-microplate hybridisation; PCR-MPH; trinucleotide repeat; TNR;
KW   SCAIII syndrome; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   repeat_region   137..355
FT   /*tag= a
FT   repeat_unit   137..139
FT   /*tag= b
FT   /note= "trinucleotide repeat"
XX
PN   WO9943852-A1.
XX
PD   02-SEP-1999.
XX
PF   18-FEB-1999; 99WO-KRO0078.
XX
PR   26-FEB-1998; 98KR-0006278.
XX
PA   (JIND/) JIN D K.
PA   (SMSU) SAMSUNG FINE CHEM CO LTD.
XX
PI   Jin DK;
XX
DR   WPI; 1999-527634/44.
XX
PT   Diagnosis of spinocerebellar ataxia type III (SCA III) syndrome
PT   using techniques which ensure highly accurate diagnosis

Claim 1: Page 12-13; 28pp; English.
XX
CC   The invention relates to the diagnosis of spinocerebellar ataxia type
CC   III (SCAIII) syndrome using reverse dot hybridisation or PCR-microplate
CC   hybridisation (PCR-MPH). The method comprises attaching a potion of the
CC   SCAIII gene containing 73 copies of the trinucleotide (CAG) repeat unit
CC   (the present sequence) to a substrate, and hybridising with amplified
CC   testee genomic DNA containing copies of the trinucleotide units, PCR
CC   amplified with labeled primers (AAX89889-90). The new method is useful
CC   for diagnosis of SCAIII syndrome, and for determining the severity of
CC   the disease. The present sequence represents the SCAIII gene fragment
CC   containing 73 trinucleotide (TNR) repeats.
XX
SQ   Sequence 397 BP; 124 A; 104 C; 104 G; 65 T; 0 other;

Query Match   Best Local Similarity   58.9%;
  Score 51.8; DB 20; Length 397;
  Matches   89; Conservative   0; Mismatches   62; Indels   0; Gaps   0;

Qy   801 GCTGCTGCTGCTGCCGCGCCGTCGTTGTTGGCTGCAACTGCGTTGTTGTTGTTTCT   860
           |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   200 GCTGCTGCTGCTGCTGTTGCTGCTGTTGCTGCTGCTGCTGCTGTTTGTTGTTGTTGCT   141

Qy   861 GCTGTAGAGAAGAAAAAGAGGGATTTCGTATTCAATTTCAAATTTGAAAAAGAGAGAGA   920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   140 GCTGTTGCTGCTTTTGCTGCTGTCTGAAAACATTCAAAAAGTGAAGTATATTTAAAAAACAA   81

Qy   921 CAAACAAAGAAAACTGAGACAGAAGAAAGTGGAAA   951
           ||| ||||||||||||||||||| ||| |||
Db   80 AACTTAAAAGAATAAAATACACCATGAGAAAA   50

RESULT 14
ABK10240/c
ID   ABK10240 standard; DNA; 403 BP.
XX
AC   ABK10240;
XX
DT   20-MAY-2002   (first entry)
XX
DE   Trinucleotide repeat sequence #3.
XX
KW   Trinucleotide repeat; fragile X syndrome; ds;
KW   spinocerebellar ataxia type III; DRPLA; neuropsychiatric;
KW   hereditary hypertrophic cardiomyopathy; Marfan syndrome;
KW   dentatorubral and pallidoluysian atrophy;
KW   spinocerebellar ataxia; X-linked spinobubar atrophy.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   repeat_region   152..347
FT   /*tag= a
FT   /rpt_type= TANDEM
FT   repeat_unit   152..154
FT   /*tag= b
FT   /note= "CAG type repeat"
XX
PN   KR200000300 4-A.
XX
PD   15-JAN-2000.
XX
PF   25-JUN-1998; 98KR-0024064.
XX
PR   25-JUN-1998; 98KR-0024064.
XX
PA   (SMSU) SAMSUNG MEDICAL CENT.
PA   (JIND/) JIN D G.
XX
PI   Jin DG;

CC   and pallidoluysian atrophy (DRPLA), spinocerebellar ataxia, Marfan
CC   syndrome; X-linked spinobubar atrophy. The present sequence
CC   represents a nucleotide sequence of the invention showing the
CC   the CAG tri-nucleotide repeats.
XX
SQ   Sequence 277 BP; 80 A; 66 C; 62 G; 69 T; 0 other;

Query Match   4.6%; Score 53.4; DB 21; Length 277;
Best Local Similarity   59.6%; Pred. No. 0.00011;
Matches   90; Conservative   0; Mismatches   61; Indels   0; Gaps   0;

Qy   801 GCTGCTCGCTGCTCGCCCGCGTCGTTGTTGTTGTGGCTGCAACTCGCTGCC 860
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   215 GCTGCTCGCTGCTCGCCCGCGTCGTTGTTGTTGTGGCTGCAACTCGCTGCC 215

Qy   861 GCTGTGAGAAGAAAAAGAGGAGATTTCGTATTCAAATTCAAAAAGAAGAGA 920
                ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   155 GCTGTTGCTGCTTGCTTTGCTCGTCTCGAAACAATTCAAAAGTGAAGTATA 156

Qy   921 CAAACAAAGAAACTGAGACACGAAAAGTGGGAAA 951
                ||| ||| || || || | |||| |||||
Db   95 AACTTAAAAGAATTAAAATACACCATGAGAAAA 65

RESULT 11
ID   ABL29756 standard; DNA; 7029 BP.
XX
AC   ABL29756;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 40741.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO20017142-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 40741; 21pp + Sequence Listing; English.
XX
XX   The invention relates to an isolated nucleic acid detection reagent
XX   capable of detecting 1000 or more genes from Drosophila. The invention is
XX   useful in developmental biology and in elucidating cell signalling and
XX   cell-cell interactions in higher eukaryotes for the development of
XX   insecticides, therapeutics and pharmaceutical drugs. The invention
XX   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7029 BP; 1426 A; 1440 C; 1945 G; 2218 T; 0 other;

Qy   801 GCTGCTCGCTGCTCGCCCGCGTCGTTGTTGTTGTGGCTGCGAACTGCTGCGCTGCCGTTCGTTGTTGTTGTTGTTGTTCT 860
                |||| ||||||||||||||||||| || |||| |||||||||||||||||||||||||||| ||||||||||||||| ||
Db   5500 GCTGTTGTTGTTGCTGCTCGCTGCTCGCCCGCGTTGTTGTTGTTGTGGCTGCGAACTGCTGCGCTGCCGTTCGTTGTTGTT 5559

Qy   861 GCTGTGAGAAGAAAAAGAGGAGATTTCGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGCT 860
                ||||||||||||||||||||||||||||||| ||||||| ||| |||| |||||||| ||||||||||||||| || ||
Db   5560 GCTGTATTTGCTAAAAAAGAGAGTAATCAATTAATTTCCGGTAAAA 5603

Query Match   4.5%; Score 52.8; DB 23; Length 7029;
Best Local Similarity   69.2%; Pred. No. 0.00086;
Matches   72; Conservative   0; Mismatches   32; Indels   0; Gaps   0;

RESULT 12
ID   ABK10248 standard; DNA; 389 BP.
XX
AC   ABK10248;
XX
DT   20-MAY-2002 (first entry)
XX
DE   Tri-nucleotide repeat related sequence #5.
XX
KW   Trinucleotide repeat; fragile X syndrome; ds;
KW   spinocerebellar ataxia type III; DRPLA; neuropsychiatric;
KW   hereditary hypertrophic cardiomyopathy; Marfan syndrome;
KW   dentatorubral and pallidoluysian atrophy;
KW   spinocerebellar ataxia; X-linked spinobubar atrophy.
XX
OS   Unidentified.
XX
FH   Key   Location/Qualifiers
FT   misc_difference 28
FT   /*tag= a
FT   /note= "illegible in the specification"
FT   misc_difference 31
FT   /*tag= b
FT   /note= "illegible in the specification"
FT   misc_difference 46
FT   /*tag= c
FT   /note= "illegible in the specification"
FT   misc_difference 56
FT   /*tag= d
FT   /note= "illegible in the specification"
FT   misc_difference 203
FT   /*tag= e
FT   /note= "illegible in the specification"
FT   misc_difference 276
FT   /*tag= f
FT   /note= "illegible in the specification"
XX
OS   Homo sapiens.
XX
PN   KR200003004-A.
XX
PD   15-JAN-2000.
XX
PF   25-JUN-1998; 98KR-0024064.
PR   25-JUN-1998; 98KR-0024064.
XX
PA   (SMSU) SAMSUNG MEDICAL CENT.
PA   (JIND/) JIN D G.
XX
PI   Jin DG;
XX
DR   WPI; 2000-662424/64.
XX
PT   Plasmid vector arraying neuropsychiatric gene more than copy number
PT   containing trinucleotide repeats its use thereof -
XX
PS   Disclosure; Fig 10; 23pp; Korean.

```
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-2145121P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75545.
XX
PS   Claim 1; Fig 6; 135pp; English.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions, and for
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of mouse B7-L_m2.
XX
SQ   Sequence 895 BP; 250 A; 223 C; 212 G; 210 T; 0 other;

Query Match           15.6%;   Score 181.8;  DB 24;  Length 895;
Best Local Similarity 72.9%;   Pred. No. 3.6e-40;
Matches  234;  Conservative  0;  Mismatches  87;  Indels  0;  Gaps  0;

Qy   42 CCGGTTCTCTGGGTCTGGTAATGAAGTCATAGAAGGCCCCAAAATGCAAGAGTCCTGAAGG 101
         |||||| |||||||||||| || |||||||||| ||||||||  |||||||| |||| ||||
Db  108 CAGCTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAAATGTAACAGTCCTAAAGG 167

Qy  102 GCTCCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCGTC 161
         ||| | ||||||| |||||||||||||| || | |||||||||||||| |||||||| ||
Db  168 ACTCAGGAGGCTCACTCAACTGCACCGTGACTCACGCGTGGAAGCTCTTCATGTGGACTC 227

Qy  162 TCAGTGACATGGTGTGCTGCTAAGGCGTCAGGCCCAATGGAGGCCCATCATCATCACCAATGACCGCT 221
         |  |||| | ||||| ||||||| | |||| |  |||||| ||| |||||||||||||  || ||||
Db  228 TTAACCAAATGGGTGGTGGTGCTGAGTCTCACCACCCAAGGACCCATCATCATCAACAACCGCT 287
```

```
Qy  222 TCACCTCTCAGAGGTACGACCAGGGCGGGGGAACTTCACCTCGGAGGATGATCATCCACAATG 281
         |||| | ||| |  | ||| ||  ||||| | ||||||||||||| | ||||| ||||| |||
Db  288 TCACCTATGCCAGTTACAAGCACACTGACAGCTTCATCTCGGAGTTGATCATCCATGATG 347

Qy  282 TGGAGCCCAGTGATTCGGGGAACATCAGATGTGCAGCCGCCGCTTCAGAAACAGTGCCCTGCTGGGAT 341
         |||| ||||||| ||  ||||| |||| ||  | ||||| || | |||||||| || || |||||
Db  348 TGCAGCCCAGTGCACTGGGGATCCGTGTCGTTCAATCCAGCCTGCAGAACAGCCAGTGGGTTTGGAT 407

Qy  342 CTGCTTACCTTACCGTCCAAG 362
         |||| | ||| ||||| ||||
Db  408 CTGCCTTCCTCTCAGTGCAAG 428
```

```
RESULT 10
XX
ID   ABK10239/c standard; DNA; 277 BP.
XX
AC   ABK10239;
XX
DT   20-MAY-2002 (first entry)
XX
DE   Trinucleotide repeat sequence #2.
XX
KW   Trinucleotide repeat; fragile X syndrome; ds; DRPLA;
KW   spinocerebellar ataxia type III; Marfan syndrome;
KW   hereditary hypertrophic cardiomyopathy; neuropsychiatric;
KW   dentatorubral and pallidolusian atrophy;
KW   spinocerebellar ataxia; X-linked spinobubar atrophy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   repeat_region   152..194
FT                   /*tag= a
FT   repeat_region   152..194
FT                   /rpt_type= TANDEM
FT   repeat_unit     152..154
FT                   /*tag= b
FT                   /note= "CAG type repeat"
FT   repeat_region   198..222
FT                   /*tag= a
FT                   /rpt_type= TANDEM
FT   repeat_unit     198..200
FT                   /*tag= b
FT                   /note= "CAG type repeat"
XX
OS   Homo sapiens.
XX
PN   KR2000003004-A.
XX
PD   15-JAN-2000.
XX
PF   25-JUN-1998;  98KR-0024064.
XX
PR   25-JUN-1998;  98KR-0024064.
XX
PA   (SMSU ) SAMSUNG MEDICAL CENT.
PA   (JIND/) JIN D G.
XX
PI   Jin DG;
XX
DR   WPI; 2000-662424/64.
XX
PT   Disclosure; Page 12; 23pp; Korean.
XX
PT   Plasmid vector arraying neuropsychiatric gene more than copy number
PT   containing trinucleotide repeats its use thereof -
XX
CC   This invention relates to the use of a plasmid vector array
CC   containing a neuropsychiatric gene containing trinucleotide repeats.
CC   Tri-nucleotide repeats and their expansion areknown to be involved
CC   in a number of diseases including fragile X syndrome, spinocerebellar
CC   ataxia type III, hereditary hypertrophic, cardiomyopathy, dentatorubral
```

RESULT 8
ID   ABK13034
XX
AC   ABK13034
XX
ID   ABK13034 standard; cDNA; 754 BP.
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding mouse B7-like protein, B7-L_m3.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Mus musculus.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   53..724
FT   /*tag= a
FT   /product= "B7-like protein, B7-L_m3"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75546.
XX
PS   Claim 1; Fig 7; 135pp; English.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ

CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of mouse B7-L_m3.
XX
SQ   Sequence 754 BP; 220 A; 191 C; 175 G; 168 T; 0 other;

Query Match   15.6%; Score 181.8; DB 24; Length 754;
Best Local Similarity   72.9%; Pred. No. 3.3e-40;
Matches 234; Conservative   0; Mismatches   87; Indels   0; Gaps   0;

Qy   42 CCGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGCCCCCAAAATGCAAGAGAGTCCTGAAGG 101
          ||| || ||| || || | ||| || |||| | ||| |||| |||||||| ||
Db   108 CAGCTTCCGGATCCAGTTATCAGATCAATAGAGAAGGTCCTCAGAATGTAACAGAGTCCTAAAGG 167

Qy   102 GCTCCCAGGCTCGCGTTCAACTGCACCGTCTCCCAGGCTGGGAAGCTCATCATGTGGGGTC 161
          ||| | |||| | | |||||| || ||| | | | |||||||||| ||||||||| |
Db   168 ACTCAGAGGCCTCACTTCAACCTGCACCGTGACTCACGGCTGGGAAGCTTCTCATGTGGACTC 227

Qy   162 TCAGTGACATGGTGGTGTGCTAAGCGTCAGGCCCCATGAACGACCCCATCATCACCACCGCT 221
          | | | ||| || || |||| || ||| | || | ||| |||||||||| | ||||
Db   228 TTAACAAAATGGTGTGGTTGCTGTGAGCGTCTCACCACCCAAGGAGACCCATCATCACCAACAACCGCT 287

Qy   222 TCACCTCTCAGAGGTACGACCAGGCGGGAACTTCACCTCGGAGATCAATCATCCACAATG 281
          ||||| | | || || | | | || || ||||| ||||| || |||||| || |||
Db   288 TCACCTTATGCCAGTTACAACAGACACTGACAGCTTCATCTCGGAGTTGATCATCCATGATG 347

Qy   282 TGGAGCCCAGTGATTGCGGGGAACATCAGATGCGAGCCTCCAGAAACAGTGCGCCTGCATGGAT 341
          || || ||||| | | ||| | | || ||| || | ||| |||| ||| | | | ||| ||||
Db   348 TGCAGCCCAGTGACTCGGGATCCGTGCAATGCAGCCTGCAGAGACAGCAGCCATGGGTTTGGAT 407

Qy   342 CTGCTTACCTTACCGTCCAAG 362
          |||| || || | | |||||
Db   408 CTGCCTTCCTCTCAGTGCAAG 428

RESULT 9
ID   ABK13033
XX
AC   ABK13033
XX
ID   ABK13033 standard; cDNA; 895 BP.
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding mouse B7-like protein, B7-L_m2.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Mus musculus.
XX
FH   Key   Location/Qualifiers
FH
FT   CDS   53..865
FT   /*tag= a
FT   /product= "B7-like protein, B7-L_m2"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.

```
Qy    42  CCGGTTCTGGGTCTGGTAATGAAGTCATGAGAAGGCCCCAAAATGCAAGAGTCCTGAAGG  101
          |||  |||||| |||||| ||  |||||||| |||||||| ||||||||||||||||||
Db   108  CAGCTTCCGGATCCAGTTATCAGATCAATAGAAGGTCCTTCAGAGATGTAACAGTCCTAAGG 167

Qy   102  GCTCCAGCGCTCGCTTCAACTGACGTCTCCGAGGCTGTGGAAGCTCATCATGTGGGCTC  161
          ||||| ||||| |||||||||| | ||| | |||| || |||||| || ||||||||| |
Db   168  ACTCAGAGGCTCACTTCAACTGCACGTGACTCACGCGTGGAAGCTTCTCATGTGGACTC  227

Qy   162  TCAGTGACATGGTGGTGCTGTAAGGGTCAGGCCCATGGAGGCCCATCACCAATGACCCGCT  221
          |||| | |   |||||||||| |  |||||| ||| |||||| |||| ||||||| ||||||
Db   228  TTAACCAAAATGGGTGGTGCTGAGTCTCACCACCACCAATCACCACCAACAACCGCT      287

Qy   222  TCACCTCTCAGAGGTACGACCAGGGGGGGAACTTCACCTCGGAGATGAATCATCCACAATG  281
          |||| | || ||||| ||| ||  || ||| ||| ||  ||| |  ||  |||||| |||
Db   288  TCACCTATGCCAGTTACAACAGACAGCACTTCATCTCGGAGTTGAACATCCATGATG      347

Qy   282  TGGAGCCCAGTGATTCGGGGAACAGAGTGACCTTCAGGACACATCAGCCGCCGCATGGAT  341
          ||||| |||||| ||||||||||| || || || ||| ||| | ||||| ||||||||||
Db   348  TCACCTATGCCAGTTACAACAGACAGCACTTCATCTCGGAGTTGAACATCCATGGAT      407

Qy   342  CTGCTTACCTTACCGTCCAAGTTATTGGAGAGCTGTGTTCATTCAATTCTTGTTAG     401
          || ||||||| || |||||| | | || |||||| || |||||| ||| |||| ||
Db   408  CTGCCTTCTCTCAGTTCAAGTCATGGGGACCCTGACAGAAACAACCTTATAG         467

Qy   402  TCGCTGAGAATGAACCTTGTGAAGTTACTGTTGCTACCCTACACTGGACCTGGCTCCCGG  461
          | ||||| | || || || |||||||||| |||||||| || |||||| ||| || |||||
Db   468  TCACTGAGGGTGAACCCTGTAATGTAGACTTGCCTGGGCTGGACCTCCACTCCCGG     527

Qy   462  ATATTCCTGGGAGCTCCGTCTCCTGGTCAGCCATTCAAAGCTATTATTTGTTCCCGGAGC  521
          |||||||||| ||| |||  ||||||||| | |||||| |||| ||  |||||| |||||
Db   528  ATATTCCTGGGAGCTTGGAGTTCCGGTTAAGCCCATTCACAGTCCAATTCCTTCTGGAGC  587

Qy   522  CCACGACCTTCAAAGTGCAGTGAGCAATCCTGCTCTCGACCCCACAGAGCAAATGGGGACTT  581
          ||| |  ||| ||| |  || || | || |||| ||| || ||||| | ||| |||| |||
Db   588  CGGGCAACTTTATGAGGGTCTTGAGTGTCCTGACCACCACTGGGCAACGGGACCT      647

Qy   582  TGACTTGCGGTGGCTACCTTGGAAGAGCCTTGAAGGCCCCGGAAGTCTGCAACTGTAAATCTCA  641
          || || |  ||||||  ||| || |  ||||| ||  ||  || |||| |||||| | |||
Db   648  TGACTTGTGTGGAGAGACTTGAAGGACTTGCAGGAAGTTCCTTAAACTGTTCAACCTGA  707

Qy   642  CTGTGATTCGGTGTCCCCAAAGACACTGGAGGTGGTATTAAATATTCCAGGTGTTGATTATCAA  701
          ||| |||| |||| | || ||||||||||||||||| ||||||| |||| ||| || |||||||
Db   708  CGTGGTTTCAGCCTCCACCTGACAGTATTGGAGAGGA-----------------------  744

Qy   702  GTTTACCGAGTTTAGGTTTTCATTGCCTACTTGGGGCAAAGTTGGACTTGGACTAGCAG  761
                  ||      ||| |  |  |      ||| || | |  |   |||  ||| |
Db   745  ---------AGGCCCAGCACTGCGACGTCTGGGCCGCCATCATCCTGCTGGCCAGTGCGCCT  791

Qy   762  GCACCAATGCTTCTGACGCCGACGTGTACTCTTACAATACGCTGCTGCTGCGCCGGTC  821
                |||| |    |||||| | | |  |||   ||||| ||| |||| ||| ||||
Db   792  TTTCCTTCGTCTCTTGATCATTGTTTGATTATAATATTCTGTTTGCTG------------  843

Qy   822  GTTGCTGTTGCGGCTGCAACTGCTGCTGCGCCGTTTGTTGTTTTCGTGTGTAGAGAAAAGAGGAT  881
                || ||   ||  |||  |  ||| |||| ||  |  ||  |||||| |  |
Db   844  ------TTGTGCCTCCAGGAGAGAAAGAGGAGAATCTACTTATCAAAATGAATAAGGAAA  898

Qy   882  TTCGTATTCAATTTCAAATAAGAGATCTGAAAAGAAGAGACAACAAAACTGAGACAG      941
          || | |  |||| | ||| |  ||  |||||  || ||||| ||||| |||||||||
Db   899  TCTGCAAACAATGAGGACAA---------ACAAAAGCAGATCGGGACAGACAAAGTTAA  944

Qy   942  AAAGTGGAAAAATGAAAAACTCCGGCTACAATTCAGATGAACAAAAGAGACCACAGCACCGCT  1001
          ||||||||||| | |||||||  ||  || || || ||  ||||||| | | || | | || |
Db   945  AAAGTGGAAAAGGAAAACTACGGGTAGCAGTTCCAGAGAAAAACAGCAGCAGTA------  1004

Qy  1002  CTCTCCCCCCAAAATCCTGTGAATCAGTTCCGATGAGCGAAAAGAAAACAGTAGCGTGTGGCC  1061
          ||||||| | |||||| | | ||||| ||||||| | | ||||||||||| |||| |  ||||
Db  1005  CTCTCCCCTCCTAAAATCGTCTGAGAAGTCAGCAGTCCACGAGACTGCAT------GCC  1058

Qy  1062  CTCCTCACCAGCCAGCAGCAAGTCACCACCAGCCACGGCTTCTT          1121
          | ||||||| ||  ||| ||||||||||||||| ||||||||||
Db  1059  TTCCTTATCAGGAACTCAAATAAACATCAGCGGACGTCCATCAACTCATCCACGGGTTCCT  1118

Qy  1122  TTAATCTGGCCAGTCCTGAGAAGGTCAGTAATTAATACAACTGTAGTTA          1167
          ||  ||| |||||||| || |||||||||||| |||| |||| ||| |||
Db  1119  TTGACATCGCCAGTCCTCAGAAGGTCAGAAAATGTGACTTTAGTGTA           1164
```

```
Query Match     16.8%;  Score 196.8;  DB 22;  Length 401;
Best Local Similarity   96.6%;  Pred. No. 1.6e-44;
Matches   201;  Conservative   0;  Mismatches   7;  Indels   0;  Gaps   0;

Qy    27  AAAATAGAGACCCACCCGGTTCTGGGTCTGGTAATGAAGTCAATAGAAGGCCCCAAAATG  86
          |||||  || ||| || |||||||||||||||||||||||||| |||||||||||||||
Db   194  AAAATTGGCTACCTTCCAGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGCCCCAAAATG  253

Qy    87  CAAGAGTCCTGAAGGGCTCCAGGCTCGCTTCAACTGCACGTCTCCCAGGCTGTGGAAGC  146
          |||||||||||||||||||||||||||||||||||||||| ||||||||||| |||||||
Db   254  CAAGAGTCCTGAAGGGCTCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGCTGGAAGC   313

Qy   147  TCATCARGTGGGCTCTCAGTGACATGGTGGTGCTGTAAGGGTCAGGCCCATGGAGCCCATCA  206
          |||||| |||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   314  TCATCATGTGGGCTCTCAGTGACATGGTGGTGCTGTAAGCGTCAGGCCCATGGAGCCCATCA  373

Qy   207  TCACCAATGACCGCTTCACCCTCTCAGAG                                 234
          |||||||||||||||||||||||||||||
Db   374  TCACCAATGACCGCTTCACCCTCTCAGAG                                 401
```

```
Db  241 AATCTTTGTAGTCGCTGAGAATGAACCTGTGTGAAGTTACTTGTCTACCCTCACACTGGACC 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  452 TGGCTCCCGGATATTTCGTCGGGAGCTCGGTCTCCTTGGTCAGCCATTCAAGCTATTATTTT 511

Db  301 CGGCTCCCCGGATAATTCCTCTGGGAGCCTCGGTCTCCTGGGTCAGCCANTTCAAGCTATTATTTT 360
        |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Qy  512 GTTCCGGAGCCCACGGACCTTCAAAGTGCAGTGAGCATCTGGCTCTCGACCCCACAGAGC 571

Db  361 GTTCCGGAGCCCACGGACCTTCAAAGTGCAGTGAGCATCTGGCTCTCGACCCCACAGAGC 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  572 AATGGGACTTTGACTTGCGTGGCTACCTGGAAGAGCCCTGAAAGTCTGCAACT 631

Db  421 AATGGGACTTTGACTTGCCGTGGCTACCTGGAAGAGCCCTGAAAGTCTGCAACT 480
        |||||||||||||||||| |||||||||||||||||||||||||||||||||
Qy  632 GTAAATCTCACTGTGATTCGGTGTCCCCAAGACACTGGAAGGGCCCGAAAGTCTGCAACT 691

Db  481 GTAAATCTCACTGTTCGGTGTCCCCAAGACACTGGAAGGGCCCGAAAGTCTGCAACT 540
        ||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Qy  692 GTATTATTCAAGTTTACCGAGTTTAGGGTTTTTCATTGCCTACTTGGGGCAAAGTTGGACTT 751

Db  541 GTATTATTCAAGTTTACCGAGTTTAGGTTTTTCATTGCCTACTTGGGGCAAAGTTGGACTT 600
        |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Qy  752 GGACTAGCAGCAGCCACCATGCTTCTGACGCGTCCACGTGTACTCTTACAATACGCTGCTGC 811

Db  601 GGACTAGCAGCAGCACCATGCTTTCTGACGCGTGTACTCTTACAATACGCTGCTGCTGC 660
        ||||||||||||| |||||||| |||||||| |||||||||||||||||||||||||
Qy  812 TGCCGCCGGTCGTGTTGTGTGGGCTGCGAAACTGCTCGTCGCCGCGTCGTGTGCTGTAGAGA 871

Db  661 TGCCCCGGTCGTGTTGTTGTGGCCGTTGTTCGTGGTTCTGCTGTTAGAACA 720
        |||| ||||||||||| || |||  |||  |||  |||  ||||||||| ||
Qy  872 AAAAGAGGATTTCGTATTC 890

Db  721 AAAAGAGGTGACAATTCTTCTGACATCTTTCCCACATCCCAGAGATTAGGTGATGCATC 780
        ||||||||| ||||
Qy  891 ------AATTTCAAAAGAAAATCTGAAAAAGAAAAGAGAAGAACAAACTGAGACAG 941

Db  781 AATGGCGTCATTTTGTATAACTCATTATGTCTGCCTTCTTGAGCCTCTCACTGTGAGA 840
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  891 AAAGTGGAAATGAAAAACTCCGGCTACAATTCAGATGAACAAAGAGCACAAACCGCGCTT 890

Db  841 CTGATCCATACCGTGCACGAAATCTGAAAAAGAGAGAAGAAACTGAGACAG 900
        |||
Qy  942 AAAGTGGAAATGAAAAACTCCGGCTACAATTCAGATGAACAAAGAGCACAAACCGCGCTT 1001

Db  901 AAAGTGGAAATGAAAAACTCCGGCTACAATTCAGATGAACAAAGAGCACAAACCGCTT 960
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1002 CTCTCCCTCCCCAAATCCTGTGAATCCAGTGATCCTGAACAAAGAAAGAAACAGTAGCGTGTGGCC 1061

Db  961 CTCTCCCTCCCCAAATCCTGTGAATCCAGTGATCCTGAACAAAGAAAGAAACAGTAGCGTGTGGCC 1020
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1062 CTCTCCACCAGCGGCGGGCCGCTGATCAACGT 1087

Db 1021 CTCTCCACCAGAGAGACAATAGACACTT 1046
        |||||||||||  ||  ||
```

RESULT 6

ABK13032

ID ABK13032 standard; cDNA; 1195 BP.
XX
AC ABK13032;
XX
DT 23-APR-2002 (first entry)
DE DNA encoding mouse B7-like protein, B7-L_m1.
XX
KW Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW reproductive disorder; graft versus host disease; autoimmune disease;
KW toxic shock syndrome; allergy; nephropathy; skin disorder;
KW endocrinopathy; lymphoproliferative disorder; gene; ss.
OS Mus musculus.
XX
FH Key   Location/Qualifiers
FT CDS   53..1165
FT   /*tag= a
FT   /product= "B7-like protein, B7-L_m1"
XX
PN WO200207710-A2.
XX
PD 03-JAN-2002.
XX
PF 28-JUN-2001; 2001WO-US20719.
XX
PR 28-JUN-2000; 2000US-214512P.
PR 28-NOV-2000; 2000US-0729264.
XX
PA (AMGE-) AMGEN INC.
XX
PI Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR WPI; 2002-130881/17.
DR P-PSDB; AAU75544.
XX
PT New B7-like polypeptides, polynucleotides and their modulators, useful
PT for diagnosing, preventing and treating reproductive, immune and
PT proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS Claim 1; Fig 5; 135pp; English.
XX
CC The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC The polypeptide, polynucleotide encoding it and antibody against (I) are
CC useful for treating B7-like polypeptide-related disease, disorders or
CC conditions including reproductive disorders (e.g. infertility,
CC miscarriage, preterm labour and delivery and endometriosis) and
CC proliferative disorders. Antibodies, soluble proteins comprising
CC extracellular domains and other regulators of B7-L polypeptides are
CC useful for enhancing the immune response to tumours. (I) plays a role in
CC growth and maintenance of cancer cells based on the observation of
CC seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC polypeptide. Hence modulators of (I) are useful for the treatment of
CC cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC in allograft transplantation, graft versus host disease, T-cell
CC dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC molecules are useful for alleviating the symptoms associated with
CC diseases involving chronic immune cell dysfunction or to treat
CC autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC purpura and psoriasis, chronic inflammatory disease such as
CC inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC are also useful as immunosuppressive agents for bone marrow and organ
CC transplantation or to prolong graft survival. B7-L molecules are also
CC useful for diagnosis and treatment of diseases involving abnormal cell
CC proliferation, including arteriosclerosis and vascular restenosis.
CC Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC shock syndrome or allosensitisation due to blood transfusions, and for
CC treatment of allergy, asthma and hypersensitivity reactions,
CC nephropathies (e.g. glomerulonephritis (Grave's disease), various
CC pemphigoid), endocrinopathies (Grave's disease), skin disorders (pemphigus and
CC (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC lymphoproliferative disorders such as multiple myeloma. The present
CC sequence represents the coding sequence of mouse B7-L_m1.
XX
SQ Sequence 1195 BP; 313 A; 304 C; 291 G; 287 T; 0 other;

Query Match   32.4%;   Score 378.4;   DB 24;   Length 1195;
Best Local Similarity   62.3%;   Pred. No. 8e-95;
Matches 701;   Conservative 0;   Mismatches 356;   Indels 69;   Gaps 4;
```

```
Db   241 GGGAACCTTCACCTCGGAGAATGATCATCCACAAATGTGGAGCGCCAGTGATTCGGGGAACATC 300
Qy   308 AGATGCAGCCTTCCAGAACACAGTTGGCCTGCCATTGGATCGTCGCTTACCTTACCGTATAG 367
         ||||||||||| |||||||||||||| ||||||| || ||||| |||||||||||||||||
Db   301 AGATGCAGCCTTCAGAACACAGTTGCCCTGCATGACATGATCTCGTCCTTACCTTATG 360
Qy   368 GGAGAGCTGTTCATTCCCAGTGTTAATCTTGTAGTGCCTGAGAATGAACCTTGTGAAGTT 427
         ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   361 GGAGAGCTGTTCATTCCCAGTGTTAATCTTGTAGTCGCTGAGAATGAACCTTGTGAAGTT 420
Qy   428 ACTTGTCTTCACCCTCACACTGGACCTGGCTCCCGGATATTTCCTTCTGGGAGCTCGTCTCTG 487
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 ACTTGTCTTCACCCTCACACTGGACCTGGCTCCCGGATATTTCCTTCTGGGAGCTCGTCTCTG 480
Qy   488 GTCAGCCAATTCAAGCTATTATTTTTGTTCCGGGAGCCCAGCCCACCCTTCAAAGTGCAGTGAGC 547
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GTCAGCCAATTCAAGCTATTATTTTTGTTCCGGGAGCCCAGCCCACCCTTCAAAGTGCAGTGAGC 540
Qy   548 ATCCTGGCTCTGACCCCACAGAGACCAAATGGGACTTTGACTTGGCTGGCTGGGCTGGGAGAGC 607
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 ATCCTGGCTCTGACCCCACAGAGACCAAATGGGACTTTGACTTGGCTGGCTGGGCTGGGAGAGC 600
Qy   608 CTGAAGGCCCGCAAGTCTGCAACTGTAAATCTCACTGTGATTCGGTGTCCCCAAGACACT 667
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 CTGAAGGCCCGCAAGTCTGCAACTGTAAATCTCACTGTGATTCGGTGTCCCCAAGACACT 660
Qy   668 GGAGGTGGTGTATTAATATTCCAGTGTATTATCAAGTTTATCCGAGTTTAGGTTTTTTCATTG 727
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GGAGGTGGTGTATTAATATTCCAGTGTATTATCAAGTTTATCCGAGTTTAGGTTTTTTCATTG 720
Qy   728 CCTACTTGGGGCAAAGTTGGACTTGGACAGGCACCACCATGCTCTTCGACGCCGACCTGT 787
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 CCTACTTGGGGCAAAGTTGGACTTGGACAGGCACCACCATGCTCTTCGACGCCGACGTGT 780
Qy   788 ACTCTTACAAATACGCTGCTGCTGCCGCGTGTGGCTGCCAACTGCTGCTGCTGC 847
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 ACTCTTACAAATACGCTGCTGCTGCCGCGTGTGGCTGCCAACTGCTGCTGCTGC 840
Qy   848 CGTTGTTGTTTCTGCTGTAGAGAAAAAGAGATTTCGTATTCAATTTCAAAAGAAAATCT 907
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 CGTTGTTGTTTCTGCTGTAGAGAAAAAGAG----------GAAATCT 878
Qy   908 GGAAAAAGAGAAGGACAGAAAACTGAGACAAAGTGAAAATGAAAAACTCCCGGCTAC 967
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   879 GAAAAAGAGAAGGAAACAAAACGAGACAAGAAGAAGTGAAAATGAAAAACTCCGGCTAC 938
Qy   968 AATTCAGATGAACAAAAGAACCACAGACACACCGCTTCTCTCCCCAAAATCCTGTGAATCC 1027
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   939 AATTCAGATGAACAAAAGGACCACAGACACCGCTTCTCTCCCCAAAATCCTGTGAATCC 998
Qy  1028 AGTGATCCTGACAAGAAGAAAAGTAGTCTGTCGTGTGGGCCCTCCTCCTCACCAGGGCGGGGCCTGATCAACGT 1087
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   999 AGTGATCCTGACAAGAAGAAAAGTAGCTGTGGCCCCTCCTCCTCACCAGGGCGGGGCTGATCAACGT 1058
Qy  1088 CCACCCAGGCCAGCAGTCATCCACAGCCTTCTTTTAATCTGGCCAGTCCTGAGAAGGTC 1147
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1059 CCACCCAGGCCAGCAGTCATCCACAGCCTTCTTTTAATCTGGCCAGTCCTGAGAAGGTC 1118
Qy  1148 AGTRATACAACTGTAGTATA 1167
         ||||||||||||||||||||
Db  1119 AGTRATACAACTGTAGTATA 1138
```

```
Query Match           66.6%;  Score 777.6;  DB 23;  Length 1392;
Best Local Similarity 87.2%;  Pred. No. 6.8e-206;
Matches  912;  Conservative   0;  Mismatches  24;  Indels  110;  Gaps  1;

Qy   152 ATGTGGGCTTCTCAGTCACATGGTGTGTGCTAAGCGTCAGGCCCATGGAGCGCCATCATCACC 211
         ||||||||||||||||| |||||||| |||||||||||||||||||||||||||||||||||
Db     1 ATGTGGGCTTCTCAGTCGACATGGTGGTCCTAAGCGTCAGGCCCATGGAGCGCCATCATCACC 60
Qy   212 AATGACCGCTTCACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATC 271
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 AATGACCGCTTCACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATC 120
Qy   272 ATCCACAAATGTGGAGCGCCAGTGATTCGGGGAACATCAGATGCAGCCTTCCAGAAACAGTCGC 331
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ATCCACAAATGTGGAGCGCCAGTGATTCGGGGAACATCAGATGCAGCCTTCCAGAAACAGTCGC 180
Qy   332 CTGCATGGAGCTTGCTTACCTTACCGTATCCCAGTTATGGGAGAGCTGTTCATTCCCAGTGTT 391
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGCATGGAGCTTGCTTACCTTACCGTATCCCAGTTATGGGAGAGCTGTTCATTCCCAGTGTT 240
Qy   392 AATCTTGTAGTCGCTGAGAATGAACCTTGTGAAGTTACTTGTCTTCACCCTCACACTGGACC 451
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy    634 AAATCTCACGTGATTCGGTCGTCTCCCCAAGACACGTGGAGGTGGTATTAATATTCCAGGTGT  693
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    706 AAATCTCACGTGATTCGGTCGTCTCCCCAAGACACGTGGAGGTGGTATTAATATTCCAGGTGT  765

Qy    694 ATTATCAAGTTTACCGAGTTTAGGTTTTTCATTGCCTACTTGGGGCAAAGTTGGACTTGG  753
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    766 ATTATCAAGTTTACCGAGTTTAGGTTTTTCATTGCCTACTTGGGGCAAAGTTGGACTTGG  825

Qy    754 ACTAGCAGGCACCATGCTTCTTGACCGCCGACGTGTACTCTTACAATACCGTGCTTGCTGCTG  813
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    826 ACTAGCAGGCACCATGCTTCTTGACCGCCGACGTGTACTCTTACAATACCGTGCTTGCTGCTG  885

Qy    814 CCGCCGTCGTGTTGTTCTCGCGTTGTTCTCGTGTAGAAGAA  873
          |||||||||||||||||||||||||||||||||||||||||
Db    886 CCGCCGTCGTTGTTGTTCTGGCTGCAACTGCGTGCCGTTGTTCTCGTGTAGAAGAA  945

Qy    874 AAGAGGATTTCGTATTCAATTTCAAAAGAAAATCTGAAAAGAAAGAGAGAACAAACAAAGAAAC  933
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    946 AAGAGGATTTCGTATTCAATTTCAAAAGAAAATCTGAAAAGAAAGAGAGAACAAACAAAGAAAC 1005

Qy    934 TGAGACAGAAAGTGAAAATGAAAACTCCCGGCTACAATTCAGATATGAACAAAAGACCACAGA  993
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1006 TGAGACAGAAAGTGGAAATGAAAACTCCCGGCTACAATTCAGATATGAACAAAAGACCACAGA 1065

Qy    994 CACCCGCTTCTCTCTCCCTCCCAAATCCTGTATCCTGATCGTCGAACAAAAGAAACAGTAG 1053
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1066 AACCCGTTCTCTCTCCCTCCCAAAATCCTGTGATCGTCGATCCTGAACCAGAAGAAACAGTAG 1125

Qy   1054 CTGTGGCCCTCCTCACCAGCGGGCGGTCTGATCAACGTCCACCCAGGCCAAGTCATCCACA 1113
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1126 CTGTGGCCCTCCTCACCAGCGGGCGGTCGATCAACGTCACCCAGGCCAAGTCATCCACA 1185

Qy   1114 GGCTTCTTTAATCTGGCCAGTCCTGAGAAGGTGCAGTAATACAACTGTGAGTATAG 1168
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1186 GGCTTCTTTAATCTGGCCAGTCCTGAGAAGGTGCAGTAATACAACTGTGAGTATAG 1240
```

RESULT 4

ABK13031

```
ID   ABK13031 standard; cDNA; 1139 BP.
XX
AC   ABK13031;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h4.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; immunosuppressive; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1134
FT                   /*tag= a
FT                   /product= "B7-like protein, B7-L_h4"
XX
PN   WO020200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
```

```
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA,  Sarmiento UM,  Schultz HJ,  Chute HT;
XX
DR   WPI; 2002-130881/17.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -.
XX
PS   Claim 1; Fig 4; 13pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h4.
XX
SQ   Sequence 1139 BP; 290 A; 300 C; 283 G; 266 T; 0 other;
```

```
Query Match             94.4%;  Score 1102.8;  DB 24;  Length 1139;
Best Local Similarity   97.9%;
Matches 1136;  Conservative    0;  Mismatches    2;  Indels   22;  Gaps  1;

Qy      8 ATGGTGGCAGGAGCCATGAAAATAGAGACCCACCCGGTTCTGGGTCTCGGTAATGAAGTC  67
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGTGGCAGGAGCCATGAAAAATAGAGACCCACCCGGTTCTGGGTCTGGTAATGAAGTC  60

Qy     68 ATAGAAGGCCCCAAAATGCAAGAAGTCCTGAAGGGCCTCGCCTTCAACTGCACC 127
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATAGAAGGCCCCAAAAATGCAAGAGTCCTGAAGGGCTCCGCCTTCAACTGCACC 120

Qy    128 GTCTCCCAGGCGTGAAGTCTCATCATGTGGGCTCTCAGTGACATGGTGTGTGTTAAGGCGTC 187
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTCTCCCAGGGCTGGAAGTCTCATCACCAAATGACCGGCTCCACATGGTGGTGTGTTAAGGCGTC 180

Qy    188 AGGCCCATGGAGCCCATCAATCACCAAATGACCGGCTTCACCTCTCAGTGACATGG 247
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGGCCCATGGAGCCCATCAATCACCAAATGACCGGCTTCACCTCTCAGAGGTACGA 240

Qy    248 GGGAACTTCACCTCGGAGATGATCATCCACAAATGTGGAGCCCAGTGATTCGGGGAACATC 307
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   944 GACAGAAAGTGGAAATGAAAACTCCGGCTACAATTCAGATCGAAGTGAAACAAAAGACCACAGACAC
Qy   997 CGCTTCTCTCCCTCCCAAAATCCTGTGAATCCAGTGAAACAAAAGAGAAACAGTAGCTG 1056
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1004 CGCTTCTCTCCCTCCCAAAATCCTGTGAATCCAGTGAAACAAAAGAGAAACAGTAGCTG 1063
Qy  1057 TGGCCCTCCTCACCGGCGGGGCTGTGATCAACGTGCCAGCAGCAAGTCATCCACAGGC
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1064 TGGCCCTCCTCACCGGCGGGGCTGTGATCAACGTGCCAGCAGCAAGTCATCCACAGGC 1116 / 1123
Qy  1117 TTCTTTAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG 1168
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1124 TTCTTTAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG 1175
```

```
RESULT 3

ID   ABK13030 standard; cDNA; 1240 BP.
XX
AC   ABK13030;
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h3.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH
FT   Key             Location/Qualifiers
FT   CDS             80..1240
FT                   /*tag= a
FT                   /product= "B7-like protein, B7-L_h3"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75542.
XX
PT   New B7-like polypeptides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 1; Fig 3; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-L polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, Pretern labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h3.
XX
SQ   Sequence 1240 BP; 318 A; 319 C; 305 G; 298 T; 0 other;
```

```
Query Match             95.8%;  Score 1119;  DB 24;  Length 1240;
Best Local Similarity   99.1%;
Matches  1125;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy    34 AGACCCACCCGGTTCTCGGTGTCTGGTAATGAAGTTCATAGAAGCCCCAAAATGCAAGAGT 93
         ||| |        |||||||||||||||||||||||||   |||||||||||||  ||||
Db   106 AGAAGCTGTAGGTTCTGGGTCTCTGGTAATGAAGTCATAGAAGAGGGCCCCAGAAATGCAACAGT 165

Qy    94 CCTGAAGGGCTCCCAGGCGTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCAT 153
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   166 CCTGAAGGGCTCCCAGGCGTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCAT 225

Qy   154 GTGGGCTCTCAGTGGACATGGTGGTGCTRAAGCGTCAGGCCCCATGGAGCCCATCATCACCAA 213
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   226 GTGGGCTCTCAGTGGACATGGTGGTGCTRAAGCGTCAGGCCCCATGGAGCCCATCATCACCAA 285

Qy   214 TGACCGGCTTCACCTCTCAGAGGTTACGACCAGGGCGGGAACTTCACCTCGGAGATGACTCAT 273
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   286 TGACCGGCTTCACCTCTCAGAGGTTACGACCAGGGCGGGAACTTCACCTCGGAGATGACTCAT 345

Qy   274 CCACAAATGTGGGAGCCCCAGTGATTCGCGGGGAACAATCAGATGCAGCCCTCCAGAACAGTCGCCT 333
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   346 CCACAAATGTGGGAGCCCCAGTGATTCGCGGGGAACAATCAGATGCAGCCCTCCAGAACAGTCGCCT 405

Qy   334 GCATGGACTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGGTTCATTCCCAGTGTTAA 393
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   406 GCATGGACTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGGTTCATTCCCAGTGTTAA 465

Qy   394 TCTTGTAGTCGCTGAGAATGAACCTTGTGAAGTTACTGTCTACCCTCACACACTGGACCTG 453
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   466 TCTTGTAGTCGCTGAGAATGAACCTTGTGAAGTTACTGTCTACCCTCACACACTGGACCTG 525

Qy   454 GCTCCCCGGATATATTCCTGGGAGACTCGGTCTCCTGGTCAGCCAGTCTAATTATTTTGT 513
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   526 GCTCCCCGGATATATTCCTGGGAGACTCGGTCTCCTGGTCAGCCAGTCTAATTATTTTGT 585

Qy   514 TCCGGAGCCCACGGACCTTCAAAGTCGCAGTGAGCCATCCTGGTCCGACCCCACACAGACAA 573
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   586 TCCGGAGCCCACGGACCTTCAAAGTCGCAGTGAGCCATCCTGGTCCGACCCCACACAGACAA 645

Qy   574 TGGGACTTTGACTTGCGGTGGCTACCTCGGAGGAGCCCGCCAAGTCTGCAACTGT 633
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   646 TGGGACTTTGACTTGCGGTGGCTACCTCGGAGGAGCCCGCCAAGTCTGCAACTGT 705
```

```
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h1.
XX
KW   Human; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FT   CDS              27..1175
FT                    /*tag= a
FT                    /product= "B7-like protein, B7-L_h1"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-2145512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75540.
XX
PS   Claim 1; Fig 1; 135pp; English.
XX
```

CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   growth and maintenance of cancer cells overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions, skin
CC   disorders (e.g. glomerulonephritis), skin disorders (pemphigus and

CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h1.
XX
SQ   Sequence 1175 BP; 295 A; 307 C; 286 G; 287 T; 0 other;

```
Query Match       96.1%;  Score 1122.4;  DB 24;  Length 1175;
Best Local Similarity   99.5%;  Pred. No. 6.8e-302;
Matches 1126;  Conservative  0;  Mismatches  6;  Indels  0;  Gaps  0;

Qy    37 CCCACCCGGTTCTGGGTCTTGCTGTGTAAGTGAAGTGCATAGAAGGCCCCAAAATGCAATGA   96
         |||| ||||||| | ||||| ||||| |||| |||||||||||||||||| |||||||||||
Db    44 CCTCCACGGTTGTCGGGTTCTGGTAATGAAGTCATAGAAGGCCCCAGAATGCAACAGTCCT  103

Qy    97 GAAGGGGCTCCCAGGCTCGCTTCAACTGCACCGGTCTCCCAGGGCCTGGAAGCTCATCATG  156
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   104 GAAGGGGCTCCCAGGCTCGCTTCAACTGCACCGGTCTCCCAGGGCCTGGAAGCTCATCGTG  163

Qy   157 GGCTCTCAGTTGACATGGTGGTTGCTAAGCGTGCAGGCCGCCCATGGAAGTCATCACCAAT  216
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   164 GGCTCTCAGTTGACATGGTGGTTGCTAAGCGTGCAGGCCGCCCATGGAAGTCATCACCAATGA  223

Qy   217 CCGCTTCAACCTCTCAGAGGTACGGACCAGGGCGGGGAACTTCACCTCGGAGATGATCATCCA  276
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   224 CCGCTTCAACCTCTCAGAGGTACGACCAGGGGGGAACTTCACCTCGGAGATGATCATCCA  283

Qy   277 CAATGTGGAGCCCAGTGATTCGGGGACATCAGATGGAGCCCTCCAGAACAGTCGCCTGCA  336
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   284 CAATGTGGAGCCCAGTGAATTCGGGGAACATCAGATGCAGCCCTCCAGAACAGTCGCCTGCA  343

Qy   337 TGGATCTGCTTACCTTACCGTCCTGGTCTCCTGGTCAGCCATTCAAGCTATTATTTTGTTCC  396
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   344 TGGATCTGCTTACCTTACCGTCCCAAGTTATGGGAGAGCGTTCATTCCCAGTGTTAATCT  403

Qy   397 TGTAGTCGCTGAGAATGAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACCTGGCT  456
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   404 TGTAGTGCGTGAGAATGAAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACCCGGGCT  463

Qy   457 CCCGGATATTTCCTGGGGAGCTCGGTCTCCTGGTCAGCCATTCAAGCTATTATTTTGTTCC  516
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   464 CCCGGATATTTCCTGGGGAGCTCGGTCTCCTGGTCAGCCATTCAAGCTATTATTTTGTTCC  523

Qy   517 GGAGCCCAGCGAACCTTCAAAAGTGCAGTGCAGCAATCCTGCTGTTGACCCCACAGAGCAATGG  576
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   524 GGAGCCCAGCGAACCTTCAAAAGTGCAGTGCAGCAATCCTGGCTCTTGAGATGATCATCCA  583

Qy   577 GACTTTGACTTGCCTGGCCTGGCTGGAAGAGAGCCCGAAGGCCCCGAAGTCTGCAACTGTAA  636
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   584 GACTTTGACTTGCCTGGCCTGGCTGGAAGAGAGCCCGAAGGCCCCGAAGTCTGCAACTGTAA  643

Qy   637 TCTCACTGTGATTCGGTGTGTCCCCAAGACACTGGAGGGTGGTAATTATTAATATTCCAGGTGTATT  696
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   644 TCTCACTGTGATTCGGTGTCCCCAAGACACTGGAGGGTGGTAATTATTAATATTCCAGGTGTATT  703

Qy   697 ATCAAGTTTACCGAGTTTAGGTTTTTTCATTGCCTTGGGGCCAAAGTTGGACTTGGACT  756
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   704 ATCAAGTTTACCGAGTTTAGGTTTTTTCATTGCCTTGGGGCCAAAGTTGGACTTGGACT  763

Qy   757 AGCAGGCACCATGCGTTCTTGACGCGCGACGTGTACTCTTACAAATAGGCTGCTGCTGCGG  816
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   764 AGCAGGCACCATGCGTTCTTGACGCGCGACGTGTACTCTTACAAATACGGCTGCTGCTGCGG  823

Qy   817 CCGTGCGTTGTTGTGTGGCGCGAACGTGCTGCGGCCGTTGTTCTCGTCGTGGTGAGAGAAAG  876
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   824 CCGTCGGTGTTGTGTGGCTGGCAACTGCTGCGTTGTTCTCTGTTTCTGCTGTGAGAGAAAG  883

Qy   877 AGGGATTTCGTATTCAAATTCAAAAGAAATCTGAAAAGAGAACAAGAAAACAAAGAAAACTGA  936
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   884 AGGGATTTCGTATTCAAATTCAAAAGAAATCTGAAAAGAGAACAAGAAAACAAAGAAAACTGA  943

Qy   937 GACAGAAAGTGGGAAATGAAAACTCCGGCTACAAATTCAGATGACAAAAAGACCACAGAAAC  996
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   996
```

```
PR   28-JUN-2000;  2000US-214512P.
PR   28-NOV-2000;  2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA,  Sarmiento UM,  Schultz HJ,  Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75541.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis .
XX
PS   Claim 1; Fig 2; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours.(I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis); vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h2.
XX
SQ   Sequence 1168 BP; 299 A; 302 C; 288 G; 279 T; 0 other;

Query Match             100.0%;  Score 1168;  DB 24;  Length 1168;
Best Local Similarity   100.0%;
Matches 1168;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy          1 AGTGATCATGGGTGGGCAGGAGGCCATGGGAAAATAGAGAGACCCACCCGGTTGTTGGGTGTCTGGTAA  60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AGTGATCATGGGTGGGCAGGAGGCCATGGGAAAATAGAGAGACCCACCCGGTTGTTGGGTGTCTGGTAA  60

Qy         61 TGAAGTCATAGAAGGAGGCCCCCAAAAATGCAAGAGAGTCCCAGGGCTCCCAGGGCTTCGGCGTTCAA  120
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TGAAGTCATAGAAGGAGGCCCCCAAAAATGCAAGAGAGTCCTCAAGGGCTCGGCGTTCGGCGTTCAA  120
```

```
Qy        121 CTGCACCGGTCTCTCCCAGGCGTGGGAGTCTCATCATGTGGGTGGTGGTGCT  180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CTGCACCGGTCTCTCCCAGGCGTGGGAGTCTCATCATGTGGGTGGTGGTGCT  180

Qy        181 AACGGTCAGGCCCCATGGAGCCCATCATCACCAAATGACCGCGCTTCACCTCTCAGAGGTACGA  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AACGGTCAGGCCCCATGGAGCCCATCATCACCAAATGACCGCGCTTCACCTCTCAGAGGTACGA  240
```

```
Db        181 AAGCGTCAGGCCCCATGGAGCCCATCATCACCAAATGACCGCGCTTCACCTCTCAGAGGTACGA  240
Qy        241 CCAGGGCGGGGAACTTCACCTCGGAGATGATCATCACACAATGTGGAGCCCAGTGATTCGGG  300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CCAGGGCGGGGAACTTCACCTCGGAGATGATCATCACACAATGTGGAGCCCAGTGATTCGGG  300
Qy        301 GAACATCAGATGCAGCCTCCAGAACAAGTCGCCTGCATGATCGTTGCTTACGCTTACCGTCCA  360
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GAACATCAGATGCAGCCTCCAGAACAAGTCGCCTGCATGATCGTTGCTTACGCTTACCGTCCA  360
Qy        361 AGTTATGGGAGGAGCTGTTCATTCCCAGTGTTTAATCTTGTAGTCGCTGCTTACTTACCGTCCA  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGTTATGGGAGGAGCTGTTCATTCCCAGTGTTTAATCTTGTAGTCGCTGCTTACTTACCGTCCA  420
Qy        421 TGAAGTTACTTGTCTACCCTCACACTGGACCTGGCCTCGGGAGGTCGG  480
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TGAAGTTACTTGTCTACCCTCACACTGGACCTGGCCTCGGGAGGTCGG  480
Qy        481 TCTCCTGGTCAGCCATTCAAGCTATTATTTGGTTCCGGAGGCCAGGACTTCAAAGTGC  540
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCTCCTGGTCAGCCATTCAAGCTATTATTTGGTTCCGGAGGCCAGGACTTCAAAGTGC  540
Qy        541 AGTGAGCATCCTGGCTCTGACCCCACAGAGACAATGGGACTTTGACTTGCGTGGCTACCTG  600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGTGAGCATCCTGGCTCTGACCCCACAGAGACAATGGGACTTTGACTTGCGTGGCTACCTG  600
Qy        601 GAAGAGCCTGAAGGCCCGCCAAGTCTGCCAACTGCAACTGTAAATCTCACTGATTCGGTCCCCA  660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GAAGAGCCTGAAGGCCCGCCAAGTCTGCCAACTGCAACTGTAAATCTCACTGATTCGGTCCCCA  660
Qy        661 AGACACTGGAGGTGGTATTAATATTCCAGGTGTGTATTATTCAAGTTTACCGAGTTTAGGTTT  720
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGACACTGGAGGTGGTATTAATATTCCAGGTGTGTATTATTCAAGTTTACCGAGTTTAGGTTT  720
Qy        721 TTCATTGCCTACTTGGGCAAAGTTGGACTTGAGCTAGCAGGACCACCATGCTTCTGACGCC  780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TTCATTGCCTACTTGGGCAAAGTTGGACTTGAGCTAGCAGGACCACCATGCTTCTGACGCC  780
Qy        781 GACGTGTACTCTTACAATACGGTGCTCCTGCTCGTTGTTGTGGCTGCAACTG  840
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GACGTGTACTCTTACAATACGGTGCTCCTGCTCGTTGTTGTGGCTGCAACTG  840
Qy        841 CTGCTGCGCGTTGTTGTTTCTGCTGTGAGAAGAAAAAGAGGATTTCGTATTCAATTTCAAAA  900
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CTGCTGCGCGTTGTTGTTTCTGCTGTGAGAAGAAAAAGAGGATTTCGTATTCAATTTCAAAA  900
Qy        901 GAAATCTGAAAAGAGAGACAAACAAAAGAAACAGTAGCTGTGGCCCCTTCCACCAGCGGGCCTGA  960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GAAATCTGAAAAGAGAGACAAACAAAAGAAACAGTAGCTGTGGCCCCTTCCACCAGCGGGCCTGA  960
Qy        961 CGGCTACAATTCAGATGAACAAAAGACCACACACGGCTTCTCTCCCTCCCCAAAATCCTG  1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 CGGCTACAATTCAGATGAACAAAAGACCACACACGGCTTCTCTCCCTCCCCAAAATCCTG  1020
Qy       1021 TGAATCCAGTGATCCTGAACAAAAGAAACAGTAGCTGTGGCCCTTCCACCAGCGGGCCTGA  1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 TGAATCCAGTGATCCTGAACAAAAGAAACAGTAGCTGTGGCCCTTCCACCAGCGGGCCTGA  1080
Qy       1081 TCAACGTCCACCCAGGCCAGCAAGTCATCCACAGGCTTCTTTAATCTGGCCAGTCCTGA  1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TCAACGTCCACCCAGGCCAGCAAGTCATCCACAGGCTTCTTTAATCTGGCCAGTCCTGA  1140
Qy       1141 GAAGGTCAGTAATACAACTGTAGTAG  1168
              ||||||||||||||||||||||||||
Db       1141 GAAGGTCAGTAATACAACTGTAGTAG  1168
```

```
RESULT 2
ABK13028
ID   ABK13028 standard; cDNA; 1175 BP.
XX
AC   ABK13028;
XX
```

```
OM nucleic - nucleic search, using sw model

                        GenCore version 5.1.4_p5.4578
                        Copyright (c) 1993 - 2003   Compugen Ltd.

Run on:        April 28, 2003, 23:25:03 ; Search time 282.628 Seconds
                        (without alignments)
                        9306.701 Million cell updates/sec

Title:         US-09-729-264-3
Perfect score: 1168
Sequence:      1 agtgatcatggtggcaggag..........gtaatacaactgtagtatag 1168

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched:      2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters:    4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     N_Geneseq_101002:*
               1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
               2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
               3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
               4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
               5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
               6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
               7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
               8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
               9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
              10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
              11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
              12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
              13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
              14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
              15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
              16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
              17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
              18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
              19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
              20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
              21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
              22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
              23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
              24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	1168	100.0	1168	24	ABK13029	DNA encoding human
2	1122.4	96.1	1175	24	ABK13028	DNA encoding human
3	1119	95.8	1240	24	ABK13030	DNA encoding human
4	1102.8	94.4	1139	24	ABK13031	DNA encoding human
5	777.6	66.6	1392	23	AAS92356	DNA encoding novel
6	378.4	32.4	1195	24	ABK13032	DNA encoding mouse
7	196.8	16.8	401	22	ABK13033	DNA encoding mouse
8	181.8	15.6	754	24	AAI36582	Probe #5268 used t
9	181.8	15.6	895	24	ABK13034	DNA encoding mouse
					ABK13033	DNA encoding mouse

ALIGNMENTS

```
RESULT 1
ABK13029
ID   ABK13029 standard; cDNA; 1168 BP.
XX
AC   ABK13029;
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h2.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             8..1168
FT                   /*tag= a
FT                   /product= "B7-like protein, B7-L_h2"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
```

```
c 10   53.4   4.6   277   21 ABK10239   Trinucleotide repe
c 11   52.8   4.5  7029   23 ABL29756   Drosophila melanog
c 12   51.8   4.4   389   21 ABK10248   Tri-nucleotide rep
c 13   51.8   4.4   397   20 AAX89891   Spinocerebellar at
c 14   51.8   4.4   403   21 ABK10240   Trinucleotide repe
c 15   49.4   4.2 46954   23 ABI16830   Drosophila melanog
c 16   48.4   4.1  3217   10 ABN91578   Rat androgen recep
c 17   48.4   4.1  3217   12 AAQ12002   Full-length rat an
c 18   48.4   4.1  4180   10 AAN91773   Rat androgen recep
c 19   47.8   4.1   381   20 ABV54466   Human prostate exp
c 20   47.6   4.1  2733   20 AAX84442   Mouse brain CNG-1
c 21   47.4   4.1  3894   23 ABL03353   Drosophila melanog
c 22   47.4   4.1 22341   23 ABL03352   Drosophila melanog
c 23   46.6   4.0  2333   22 AAF75339   Human TGF-beta rec
c 24   46.4   4.0  1983   23 ABI24483   Drosophila melanog
c 25   46.4   4.0  4044   23 ABI24482   Drosophila melanog
c 26   45.8   3.9  2682   23 ABI16097   Drosophila melanog
c 27   45.8   3.9  8821   23 ABI16096   Drosophila melanog
c 28   45.8   3.9  1509   23 ABI24571   Drosophila melanog
c 29   45.6   3.9  3884   23 ABI24570   Drosophila melanog
c 30   45.4   3.9  3042   23 ABI28445   Drosophila melanog
c 31   45.4   3.9  3135   23 ABI28444   Drosophila melanog
c 32   45.4   3.9  5215   23 ABI25654   Drosophila melanog
c 33   45.4   3.9  5447   23 ABI25653   Drosophila melanog
c 34   45.4   3.9  5397   23 ABI25652   Drosophila melanog
c 35   45.2   3.9   462   23 ABI29757   Drosophila melanog
c 36   45     3.9  2115   23 ABI26941   Drosophila melanog
c 37   45     3.9  2260   23 ABI26940   Drosophila melanog
c 38   45     3.9  3210   23 ABL03191   Drosophila melanog
c 39   45     3.9  3624   23 ABL08977   Drosophila melanog
c 40   45     3.9  5975   23 ABL03190   Drosophila melanog
c 41   45     3.9  6541   23 ABL08976   Drosophila melanog
c 42   44.8   3.8   254   21 ABL04172   Drosophila melanog
c 43   44.8   3.8  1310   20 AAZ45758   Polymorphic region
c 44   44.8   3.8  2614   24 AAV08558   Transgene for epit
c 45   44.8   3.8  3263   20 AAV08560   Transgene for epit
                   2614   24 ABK36100   cDNA sequence #491
                                        Transgene for epit
```

```
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 29
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 33
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 21
US-09-864-761-2182

Query Match             3.7%;   Score 43.2;   DB 10;   Length 459;
Best Local Similarity   79.7%;   Pred. No. 0.0075;
Matches   51;   Conservative   0;   Mismatches   13;   Indels   0;   Gaps   0;

Qy   801 GCTGCTGCTGCTGCCCGGGTCGCGTTGGTTGTTGGGCTGCAAACTGCTGCCGTTGTTGTTTCT 860
         ||||||||||||||| ||| ||| || ||||| |||||| |||||||||| ||||||||||| ||
Db   389 GCTGCTGCTGCCTGTTGTTGCTGCTCGTGCTGCTGCTGTTGCTGCTGCTGCTGTTGTTGCT 330

Qy   861 GCTG 864
         ||||
Db   329 GCTG 326
```

RESULT 15
US-09-416-384A-6/c
; Sequence 6, Application US/09416384A
; Patent No. US20020081584A1
; GENERAL INFORMATION:
;   APPLICANT: BLUMENFELD, Marta
;   APPLICANT: BOUGUELERET, Lydie
;   APPLICANT: CHUMAKOV, Ilya
;   APPLICANT: COHEN, Daniel
;   APPLICANT: ESSIOUX, Laurent
;   TITLE OF INVENTION: Genes, proteins and biallelic markers related to central...
;   FILE REFERENCE: GENSET.045AUS
;   CURRENT APPLICATION NUMBER: US/09/416,384A
;   CURRENT FILING DATE: 1999-10-12
;   PRIOR APPLICATION NUMBER: 60/106,457
;   PRIOR FILING DATE: 1999-10-30
;   PRIOR APPLICATION NUMBER: 60/103,955
;   PRIOR FILING DATE: 1998-10-12
;   PRIOR APPLICATION NUMBER: 60/132,277
;   PRIOR FILING DATE: 1999-05-03
;   NUMBER OF SEQ ID NOS: 71
;   SOFTWARE: Patent.pm
;   SEQ ID NO 6
;   LENGTH: 1791
;   TYPE: DNA
;   ORGANISM: mus musculus
US-09-416-384A-6

Query Match   3.7%;   Score 43.2;   DB 10;   Length 1791;
Best Local Similarity   75.0%;   Pred. No. 0.018;
Matches   54;   Conservative   0;   Mismatches   18;   Indels   0;   Gaps   0;

```
Qy   800 CGCTGCTGCTGCTGCCCGCGCGTCGTTGTTGGGCTGCAACTGCGTTGCCGTTGTTGTTTC 859
         ||||||||||||||||| ||| |||||||||||||| |||||| ||||||||||||||||
Db   336 CGCTGCTGCTGCTGCTGCTGTTGCTGCTGCTGTTGCTGCTGCTGCTGTTGCTGCTGCTGC 277

Qy   860 TGCTGTAGAAGA 871
         |||| ||| ||
Db   276 TGCTGTTGGTGA 265
```

Search completed: April 28, 2003, 21:07:44
Job time : 2026.54 secs

Db 184 GCTGCTNGCTNGCTGTTNGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGTTGTTGTTGTTGCT 125
         ||||||||||||||||| ||||||||||||||||||| ||||||||||||||||||||| ||| || ||
Qy 861 GCTG 864
         ||||
Db 124 GCTG 121

RESULT 14
US-09-864-761-2182/c
Sequence 2182, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
;   APPLICANT: Penn, Sharron G.
;   APPLICANT: Rank, David R.
;   APPLICANT: Hanzel, David K.
;   APPLICANT: Chen, Wensheng
;   TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
;   TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;   FILE REFERENCE: Aeomica-X-1
;   CURRENT APPLICATION NUMBER: US/09/864,761
;   CURRENT FILING DATE: 2001-05-23
;   PRIOR APPLICATION NUMBER: US 60/180,312
;   PRIOR FILING DATE: 2000-02-04
;   PRIOR APPLICATION NUMBER: US 60/207,456
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: US 09/632,366
;   PRIOR FILING DATE: 2000-08-03
;   PRIOR APPLICATION NUMBER: GB 24263.6
;   PRIOR FILING DATE: 2000-10-04
;   PRIOR APPLICATION NUMBER: US 60/236,359
;   PRIOR FILING DATE: 2000-09-27
;   PRIOR APPLICATION NUMBER: PCT/US01/00666
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00667
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00664
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00662
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00663
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00668
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00665
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00669
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00661
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00670
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: US 60/234,687
;   PRIOR FILING DATE: 2000-09-21
;   PRIOR APPLICATION NUMBER: US 09/608,408
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: US 09/774,203
;   PRIOR FILING DATE: 2001-01-29
;   NUMBER OF SEQ ID NOS: 49117
;   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;   SEQ ID NO 2182
;   LENGTH: 459
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC009954.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 47
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 53
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 69
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 27
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 16
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 21
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 45

;   APPLICANT: Penn, Sharron G.
;   APPLICANT: Rank, David R.
;   APPLICANT: Hanzel, David K.
;   APPLICANT: Chen, Wensheng
;   TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;   TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;   FILE REFERENCE: Aeomica-X-1
;   CURRENT APPLICATION NUMBER: US/09/864,761
;   CURRENT FILING DATE: 2001-05-23
;   PRIOR APPLICATION NUMBER: US 60/180,312
;   PRIOR FILING DATE: 2000-02-04
;   PRIOR APPLICATION NUMBER: US 60/207,456
;   PRIOR FILING DATE: 2000-05-26
;   PRIOR APPLICATION NUMBER: US 09/632,366
;   PRIOR FILING DATE: 2000-08-03
;   PRIOR APPLICATION NUMBER: GB 24263.6
;   PRIOR FILING DATE: 2000-10-04
;   PRIOR APPLICATION NUMBER: US 60/236,359
;   PRIOR FILING DATE: 2000-09-27
;   PRIOR APPLICATION NUMBER: PCT/US01/00666
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00667
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00664
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00662
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00663
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00668
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00665
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00669
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00661
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: PCT/US01/00670
;   PRIOR FILING DATE: 2001-01-30
;   PRIOR APPLICATION NUMBER: US 60/234,687
;   PRIOR FILING DATE: 2000-09-21
;   PRIOR APPLICATION NUMBER: US 09/608,408
;   PRIOR FILING DATE: 2000-06-30
;   PRIOR APPLICATION NUMBER: US 09/774,203
;   PRIOR FILING DATE: 2001-01-29
;   NUMBER OF SEQ ID NOS: 49117
;   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;   SEQ ID NO: 18923
;   LENGTH: 293
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC009954.1
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 47
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 53
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 69
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 27
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 16
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 21
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 45
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 29
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 33
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 21
;   OTHER INFORMATION: NT HIT: Z22828.1, EVALUE 9.00e-59
;   OTHER INFORMATION: SWISSPROT HIT: P53360, EVALUE 4.00e-15
;   OTHER INFORMATION: EST_HUMAN HIT: AW409956.1, EVALUE 1.00e-58
US-09-864-761-18923

Query Match   3.7%;   Score 43.2;   DB 10;   Length 293;
Best Local Similarity   79.7%;   Pred. No. 0.0056;
Matches 51;   Conservative 0;   Mismatches 13;   Indels 0;   Gaps 0;

Qy 801 GCTGCTGCTGCTCCGCGCGTCGTTGTTGTTGTTGGCTGCAACTGCTGCCTGCGTTGTTGTTTCT 860

Qy   801 GCTGCTGCTGCTGCCCCGTCGTGTTGTTGTGGCTGCAAACTGCTGCTGCCCGTTGTTGTTTCT 860
          ||| ||| ||||||||| ||||| ||||||| |||||||| ||||||||||||||||||||| ||
Db   6175 GCTGTTGCTGCTGCTGCTCGTGTTGCTGCCTGCTGCTGCTGCTGCTGTTGCTGTTGTTTGCT 6116

Qy   861 GCTGTGAAGAA 872
          ||||| | |||
Db   6115 GCTGCTGCCGCA 6104

; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 544
; TYPE: DNA
; LENGTH: 384
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-544

RESULT 10
US-09-728-444-151
; Sequence 151, Application US/09728444
; Patent No. US20020161207A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US2002016120702A1e1 Murine Polynucleotide Sequences
; TITLE OF INVENTION: and Mutant Cells and Mutant Animals Defined Thereby
; FILE REFERENCE: LEX-0100-USA
; CURRENT APPLICATION NUMBER: US/09/728,444
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 2000-11-30
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 1206
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 151
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(455)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-444-151

Query Match   3.8%;   Score 44;   DB 9;   Length 455;
Best Local Similarity   69.4%;   Conservative   0;   Mismatches   26;   Indels   0;   Gaps   0;
Matches   59;

Qy   803 TGCTGCTGCTGCCGCGTCGTTGTTGTTGGCGTGCAACTGCTGCTGCCGTTGTTGTTTCTGC 862
          ||||||||| ||| ||||||||| || | |||||| | |||| |||| ||| | |||||||
Db   162 TGCTGCTGCTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG | || | |||||| GC 221

Qy   863 TGTGAAGAAAAAAGAGGATTTCGTA 887
          ||-|| ||||| |||||| |||| |
Db   222 TGCAAACCTANACCAGATGTGGGA 246

Query Match   3.7%;   Score 43.4;   DB 9;   Length 384;
Best Local Similarity   64.4%;   Conservative   0;   Mismatches   36;   Indels   0;   Gaps   0;
Matches   65;

Qy   799 ACGCTGCTGCTGCCGCGTCGTTGTTGTTGGCGTGCAACTGCTGCTGCCGTTGTTGTTT 858
          ||-|| |||||| ||-| ||||||| || | |||||| |||||| |||| | |||||| ||
Db   98 ACGTAACCGCAGCTGCTCCTCCAGTTGCAGTTGCCTGCTGCTGCAGGCGCGTGCAGCGCTG 157

Qy   859 CTGCTGTAGAAGAAAAAAGAGGATTTCGTATTCAATTTCAAA 899
          ||-|||| | |||| |||||||||||| || || || ||||
Db   158 CTGCTGCAGAAGAGAAGGACGAGTTCGACGTCGTTCTCGAA 198

RESULT 12
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match   3.7%;   Score 43.4;   DB 9;   Length 3309400;
Best Local Similarity   64.4%;   Conservative   0;   Mismatches   36;   Indels   0;   Gaps   0;
Matches   65;

Qy   799 ACGCTGCTGCTGCCGCGTCGTTGTTGTTGGCGTGCAACTGCTGCTGCCGTTGTTGTTT 858
          ||-|| |||||| ||-| ||||||| || | |||||| |||||| |||| | |||||| ||
Db 510688 ACGTAACCGCAGCTGCTCCTCCAGTTGCAGTTGCCTGCTGCTGCAGGCGCGTGCAGCGCTG 510747

Qy   859 CTGCTGTAGAAGAAAAAAGAGGATTTCGTATTCAATTTCAAA 899
          ||-|||| | |||| |||||||||||| || || || ||||
Db 510748 CTGCTGCAGAAGAGAAGGACGAGTTCGACGTCGTTCTCGAA 510788

RESULT 13
US-09-864-761-18923/C
; Sequence 18923, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:

Meisterernst, Michael
Polites, Greg
TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSCRIPTION
COMPLEXES FROM TRANSGENIC
NON-HUMAN ANIMALS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: HELLER, EHRMAN, WHITE & McAULIFFE
STREET: 1666 K Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/849,243
FILING DATE: 07-May-2001
ATTORNEY/AGENT INFORMATION:
NAME: Granados, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 38005-0148
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)912-2000
TELEFAX: (202)912-2020
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 3263 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: exon
LOCATION: 1..3263
SEQUENCE DESCRIPTION: SEQ ID NO: 15:

US-09-849-243-15

Query Match   3.8%;   Score 44.8;   DB 9;   Length 3263;
Best Local Similarity   81.2%;   Pred. No. 0.0087;
Matches 52;   Conservative 0;   Mismatches 12;   Indels 0;   Gaps 0;

Qy   801   GCTGCTCGTCGCTGCGCCGCCCGTCGTTGTTGGCTGCGAACTGCTGCCGCCGTTGTTGTTTCT   860
            |||||||||| ||| ||||| ||||| ||||||||| ||||||||||||||||||||| ||
Db   1430   GCTGCTGCTGCTGCTGCTTCCTGTTGCTCCTGCTGCTGCTGTTGTTGTTTGTTTGTTTGCT   1371

Qy   861   GCTG   864
            ||||
Db   1370   GCTG   1367

RESULT 8
US-09-849-243-14/c
Sequence 14, Application US/0984243
Patent No. US20020157127A1
GENERAL INFORMATION:
APPLICANT: Kirschbaum, Bernd
Berglund, Erick
Meisterernst, Michael
Polites, Greg
TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSCRIPTION
COMPLEXES FROM TRANSGENIC
NON-HUMAN ANIMALS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: HELLER, EHRMAN, WHITE & McAULIFFE
STREET: 1666 K Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA

ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/849,243
FILING DATE: 07-May-2001
ATTORNEY/AGENT INFORMATION:
NAME: Granados, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 38005-0148
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)912-2000
TELEFAX: (202)912-2020
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4286 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: exon
LOCATION: 1..4286
SEQUENCE DESCRIPTION: SEQ ID NO: 14:

US-09-849-243-14

Query Match   3.8%;   Score 44.8;   DB 9;   Length 4286;
Best Local Similarity   81.2%;   Pred. No. 0.01;
Matches 52;   Conservative 0;   Mismatches 12;   Indels 0;   Gaps 0;

Qy   801   GCTGCTGCTGCTGCGCCGCCCGTCGTTGTTGTGGCTGCGAACTGCTGCCGCCGTTGTTGTTTCT   860
            |||||||||| ||| ||||| ||||| |||||||| |||||||||||||||||||||||| ||
Db   2767   GCTGCTGCTGCTGCTGCTTCCTGTTGCTGCTGCTGCTGCTGTTGTTGTTTGTTTGTTTGCT   2708

Qy   861   GCTG   864
            |||
Db   2707   GCTG   2704

RESULT 9
US-09-880-107-1748/c
Sequence 1748, Application US/09880107
Patent No. US20020142981A1
GENERAL INFORMATION:
APPLICANT: Horne, Darci T.
APPLICANT: Vockley, Joseph G.
APPLICANT: Scherf, Uwe
APPLICANT: Gene Logic, Inc.
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
FILE REFERENCE: 44921-5028:WO
CURRENT APPLICATION NUMBER: US/09/880,107
CURRENT FILING DATE: 2001-06-14
PRIOR APPLICATION NUMBER: US 60/211,379
PRIOR FILING DATE: 2000-06-14
PRIOR APPLICATION NUMBER: US 60/237,054
PRIOR FILING DATE: 2000-10-02
NUMBER OF SEQ ID NOS: 3950
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1748
LENGTH: 6604
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 D83783
US-09-880-107-1748

Query Match   3.8%;   Score 44.8;   DB 10;   Length 6604;
Best Local Similarity   76.4%;   Pred. No. 0.014;
Matches 55;   Conservative 0;   Mismatches 17;   Indels 0;   Gaps 0;

APPLICANT: Owen, Thomas A.
APPLICANT: Smock, Steven L.
TITLE OF INVENTION: The Canine Androgen Receptor
FILE REFERENCE: PC10893AGPR
CURRENT APPLICATION NUMBER: US/10/008,739A
CURRENT FILING DATE: 2002-04-15
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin version 3.1

SEQ ID NO 1
LENGTH: 3577
TYPE: DNA
ORGANISM: Canine
US-10-008-739A-1

RESULT 5
US-09-849-243-13/c
Sequence 13, Application US/09849243
Patent No. US20020157127A1
GENERAL INFORMATION:
APPLICANT: Kirschbaum, Bernd
         Berglund, Erick
         Meistererrnst, Michael
         Polites, Greg
TITLE OF INVENTION: PURIFICATION OF HIGHER ORDER TRANSCRIPTION
         COMPLEXES FROM TRANSGENIC
         NON-HUMAN ANIMALS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: HELLER, EHRMAN, WHITE & McAULIFFE
STREET: 1666 K Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/849,243
FILING DATE: 07-May-2001
ATTORNEY/AGENT INFORMATION:
NAME: Granados, Patricia D.
REGISTRATION NUMBER: 33,683
REFERENCE/DOCKET NUMBER: 38005-0148
TELECOMMUNCATION INFORMATION:
TELEPHONE: (202)912-2000
TELEFAX: (202)912-2020
INFORMATION FOR SEQ ID NO 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1310 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: exon
LOCATION: 1.1310
SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Query Match   3.9%; Score 46; DB 9; Length 3577;
Best Local Similarity   65.7%; Pred. No. 0.004;
Matches 67; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 769 GCTTCTGACGCGACGTGTACTTCTTACAAATACGCTCCTTGCTGCCGCGCCGTCGTTGTTG 828
        ||| ||| | ||| |||||||||||||||||| || | |||| | ||||| ||
Db 633 .GCTGTGCTACCTTCTTGATACTACCTCCTGCTGTTGCCTGCTGCCTGCTTGGTGGTTG 574

Qy 829 TGGCTGCCAACTGCTGCTGCCGCCTTGTTGTTTTCTGCGTGTAGAAG 870
        ||||| ||||||||||||| | || ||| || |||||| || | ||
Db 573 TTGCTGCTGCGTGCTGCTGCCGCCGCTGCTGCTGCTGAAGAG 532

RESULT 6
US-09-822-846-13

Query Match   3.8%; Score 44.8; DB 9; Length 1310;
Best Local Similarity   81.2%; Pred. No. 0.0048;
Matches 52; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 801 GCTGCTGCTGCTGCGGCGCCGTCGTTGTTGTTGGGCTGCGCAAACTGCTGCCGTTGTTGTTTCT 860
Db 336 GCTGCTGCTGCTGCTG... GTTGCTGGCTGCTGCTGCTGTTGCTGTTGTTGTTGGCT 277

Qy 861 GCTG 864
        ||||
Db 276 GCTG 273

RESULT 6
US-09-822-846-491/c
Sequence 491, Application US/09822846
Publication No. US2003027139A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: LaVallie, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steininger II, Robert J.
APPLICANT: Bowman, Michael R.
APPLICANT: Spaulding, Vikki
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakar
APPLICANT: Graham, James R.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6400
CURRENT APPLICATION NUMBER: US/09/822,846
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195,605
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 629
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 491
LENGTH: 2614
TYPE: DNA
ORGANISM: Homo sapiens
US-09-822-846-491

Query Match   3.8%; Score 44.8; DB 9; Length 2614;
Best Local Similarity   76.4%; Pred. No. 0.0075;
Matches 55; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 801 GCTGCTGCTGCTGCGGCGCCGTCGTTGTTGTTGGGCTGCGGAACTGCTGCCGTTGTTGTTTCT 860
Db 2179 GCTGTTGCTGCTGCTGCTGCTGTTGCTGTTGCTGCTGCTGCTGTTGTTGTTGCTGCTTGCT 2120

Qy 861 GCTGTGAAGANA 872
        ||||| | | |
Db 2119 GCTGCTGCCGCA 2108

RESULT 7
US-09-849-243-15/c
Sequence 15, Application US/09849243
Patent No. US20020157127A1
GENERAL INFORMATION:
APPLICANT: Kirschbaum, Bernd
         Berglund, Erick

```
Qy   234 GCGGGCTGATCAAGCTCCACCCAGCCGCGAAGTCATCCACAGCGTTCTTTTAATCTGGC 293
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1132 CAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG 1168

Qy   294 CAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG 330
         |||||||||||||||||||||||||||||||||||||
Db   294 CAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG 330

RESULT 2
US-09-864-761-16305
; Sequence 16305, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT: Rank, David R.
;  APPLICANT: Hanzel, David K.
;  APPLICANT: Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US20020048763A1
;  CURRENT FILING DATE: 2001-05-23
;  PRIOR APPLICATION NUMBER: US/09/864,761
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 60/232,366
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 16305
;  LENGTH: 401
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  OTHER INFORMATION: MAP TO AF121782.1
;  OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
US-09-864-761-16305

Query Match             16.8%;  Score 196.8;  DB 10;  Length 401;
Best Local Similarity   96.6%;  Pred. No. 6.5e-50;
Matches  201; Conservative   0; Mismatches   7; Indels   0; Gaps   0;
```

```
Qy    27 AAAATAGAGACACCCACCCGGTTCTCGGGTCTGGGTGGTAATGAAGTCATAGAAGGCCCCCCCAAAATG 86
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194 AAAATGGGCTACCTTCCAGGTTCTTCTGGGTCTGGGTGGTAATGAAGTCATAGAAGGCCCCCCCAAAATG 253

Qy    87 CAAGAGTCCTGAAGGCCTCCCAGGCCTCGGCTTCAACTGCACCGTTCTCCCAGGGGTGGAAGC 146
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314 TCATCATGTGGGCTTCTCAGTGACATGGTTGGTGCTAAGCGTCAGGCCCATTGGGAGCCATCA 373

Qy   254 CAAGAGTCCTGAAGGGCTCCCAGGCCTCGGCTTCAACTGCACCGTTCTCCCAGGGGTGGAAGC 313
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   374 TCACCAAATGACCCGCTTCACCTCTCAGAG 401

Qy   147 TCATCATGTGGGCTTCTCAGTGACATGGTTGGTGCTAAGCGTCAGGCCCATTGGGAGCCATCA 206
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   314 TCATCATGTGGGCTTCTCAGTGACATGGTTGGTGCTAAGCGTCAGGCCCATTGGGAGCCATCA 373

Qy   207 TCACCAAATGACCCGCTTCACCTCTCAGAG 234
         ||||||||||||||||||||||||||||||
Db   374 TCACCAAATGACCCGCTTCACCTCTCAGAG 401

RESULT 3
US-09-983-965-4945
; Sequence 4945, Application US/09983965
; Patent No. US20020137160A1
; GENERAL INFORMATION:
;  APPLICANT: Warren, Wesley C.
;  APPLICANT: Tao, Nengbing
;  APPLICANT: Byatt, John C.
;  APPLICANT: Mathialagan, Nagappan
;  TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;  TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;  FILE REFERENCE: 37-21(10297)C
;  CURRENT APPLICATION NUMBER: US/09/983,965
;  CURRENT FILING DATE: 2001-10-26
;  PRIOR APPLICATION NUMBER: US 09/465,231
;  PRIOR FILING DATE: 1999-12-15
;  PRIOR APPLICATION NUMBER: US 60/113,678
;  PRIOR FILING DATE: 1998-12-17
;  NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 4945
;  LENGTH: 398
;  TYPE: DNA
;  ORGANISM: Bos taurus
;  FEATURE:
;  OTHER INFORMATION: Clone ID: 26-LIB34-017-Q1-E1-G9
US-09-983-965-4945

Query Match             11.7%;  Score 136.2;  DB 10;  Length 398;
Best Local Similarity   77.5%;  Pred. No. 2.3e-31;
Matches  165; Conservative   0; Mismatches  48; Indels   0; Gaps   0;
```

```
Qy    42 CCGGTTCTGGGTGTTGGGTTAATGAAGTCATAGAAGGCCCCCAAAATGCAAGAGTCCTGAAGG 101
         ||| ||||||| ||   | ||||||||||||||||||| | | |||||  |  | |||||||||
Db   186 CAGCCTGTGGATCCACAGTGAAATCATAGAGGTCCCAAGAAGAATCTACAGACGCCCTGAAGG 245

Qy   102 GCTCCCAGGCTCGGCTTCAACTGCACCGTTCTCCCAGGGCTGGAAGTCATCATGTGGGCTC 161
         || |  | | | ||| | |||   ||||| ||| |  |||||| ||| ||| || ||||| |
Db   246 GCTCGGAGGCTCGCTTCCAACTGCACCATCTCGCACGGGCTGGAAGCTCGTCATGTGGGCTC 305

Qy   162 TCAGTGACATGGTGGTGCTAAGCGTCAGGCCCATGGGAGCCCATCACAAATGACCCGCT 221
         ||| | || | ||| ||||||||  |  | | ||| || ||| || |  ||||| ||||
Db   306 TGAGAGGCACAGTGGTGCTGAGCATGACACCTAATGAGACCATCACCAGTGACGACCGCT 365

Qy   222 TCACCCTCTCAGAGGTACGACCAGGGCGGGGGAACT 254
         ||| ||| ||| || || ||| |||| | |||||||
Db   366 TCACTTCGGCAAGCTACCAAGAGGGCCGGGAACT 398

RESULT 4
US-10-008-739A-1/c
; Sequence 1, Application US/10008739A
; Patent No. US20020161194A1
; GENERAL INFORMATION:
;  APPLICANT: Pfizer Inc.
;  APPLICANT: Castleberry, Tessa A.
;  APPLICANT: Lu, Bihong
```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 28, 2003, 16:24:32 ; Search time 138.543 Seconds
                  (without alignments)
                  9173.565 Million cell updates/sec

Title: US-09-729-264-3
Perfect score: 1168
Sequence: 1 agtgatcatggtggcaggag..........gtaatacaactgtagtatag 1168

Scoring table: IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0
Searched: 709820 seqs, 544064369 residues
Total number of hits satisfying chosen parameters: 1419640

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCI_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB	ID	Description
1	269	474	10	US-09-918-995-3342	Sequence 3342, Ap
2	196.8	401	10	US-09-864-761-16305	Sequence 16305, A
3	136.2	398	10	US-09-983-965-4945	Sequence 4945, Ap
4	46	577	9	US-10-008-739A-1	Sequence 1, Appli
5	44.8	1310	9	US-09-849-243-13	Sequence 13, Appl
6	44.8	2614	9	US-09-822-846-491	Sequence 491, Appl
7	44.8	3263	9	US-09-849-243-15	Sequence 15, Appl
8	44.8	4286	9	US-09-849-243-14	Sequence 14, Appl
9	44.8	6604	10	US-09-880-107-1748	Sequence 1748, Ap
10	44	455	9	US-09-728-444-151	Sequence 151, App
11	43.4	384	9	US-09-738-626-544	Sequence 544, Appl
12	43.4	3309400	9	US-09-738-626-1	Sequence 1, Appli
13	43.2	293	10	US-09-864-761-18923	Sequence 18923, A
14	43.2	459	10	US-09-864-761-2182	Sequence 2182, Ap
15	43.2	1791	10	US-09-416-384A-6	Sequence 6, Appli
16	42.6	2424	9	US-10-029-217A-25	Sequence 25, Appl
17	42.6	4959	9	US-10-029-217A-1	Sequence 1, Appli
18	42.6	4960	9	US-10-029-217A-30	Sequence 30, Appl
19	42.2	5959	10	US-09-954-456-1996	Sequence 1996, Ap

c	20	41.6	238	10	US-09-864-761-24079	Sequence 24079, A
c	21	41.6	554	10	US-09-864-761-7357	Sequence 7357, Ap
c	22	41.6	3899	10	US-09-735-367B-5	Sequence 5, Appli
c	23	41.6	6504	10	US-09-969-347-171	Sequence 171, App
c	24	41.6	6504	10	US-09-735-367B-9	Sequence 9, Appli
c	25	41.6	6878	10	US-09-735-367B-1	Sequence 1, Appli
c	26	41.4	2572	10	US-09-925-300-486	Sequence 486, App
c	27	41.2	424	9	US-09-918-995-10212	Sequence 10212, A
c	28	41.2	725	10	US-09-919-580-878	Sequence 878, App
c	29	41	2517	10	US-09-476-242-16	Sequence 16, Appl
c	30	40.8	2352	10	US-09-476-242-26	Sequence 26, Appl
c	31	40.8	2536	9	US-09-822-846-129	Sequence 129, App
c	32	40.8	6177	10	US-09-916-145-1	Sequence 1, Appli
c	33	40.8	6322	10	US-09-917-800A-1546	Sequence 1546, Ap
c	34	40.6	547	9	US-09-918-995-11012	Sequence 11012, A
c	35	40.4	954	10	US-10-003-295-1	Sequence 1, Appli
c	36	40.4	2674	9	US-10-078-090-83	Sequence 83, Appl
c	37	40.4	216	9	US-09-728-444-25	Sequence 25, Appl
c	38	40.2	545	10	US-09-919-580-371	Sequence 371, App
c	39	40.2	557	10	US-09-919-580-475	Sequence 475, App
c	40	40.2	572	10	US-09-919-580-570	Sequence 570, App
c	41	40.2	1143	9	US-09-931-842A-1399	Sequence 1399, App
c	42	40.2	2523	10	US-09-475-242-15	Sequence 15, Appl
c	43	40.2	83450	9	US-09-811-469-3	Sequence 3, Appli
c	44	40	270	9	US-10-007-557-8	Sequence 8, Appli
c	45	40	488	9	US-09-918-995-19918	Sequence 19918, A

ALIGNMENTS

RESULT 1
US-09-918-995-3342
Sequence 3342, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SEQ ID NO 3342
SOFTWARE: FastSEQ for Windows Version 3.0
LENGTH: 474
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(474)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-3342

Query Match   23.0%; Score 269; DB 9; Length 474;
Best Local Similarity   98.2%; Pred. No. 5.8e-72;
Matches 272; Conservative   0; Mismatches   5; Indels   0; Gaps   0;

Qy   892 ATTTCAAAAGAAATCTGAAAAAGAGCAAACAAAAGAAAACTGAGACAGAAAGTGGAAA 951
            ||| ||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db   54 ACTTTTAAGGAAAATCTGAAAAAGAGAGAAGAGACAAACAAAAACTGAGACAGAAAGTGGAAA 113

Qy   952 TGAAAACTCCGGCTACAAATTCAGATGAACAAAAGACCACAGACACCGCTTCTCCCTCC 1011
            |||||| |||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   114 TGAAAACTCCGGCTACAAATTCAGATGAACAAAAGACCACAGAAACCGCTTCTCTCCCTCC 173

Qy   1012 CAAATCCTGTGAATCCAGTGATCCTGAACAAAAGAAAACAGTAGCGTGGCCCTCCTCACCA 1071
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   114 CAAATCCTGTGAATCCAGTGATCCTGAACAAAAGAAAACAGTAGCGTGGCCCTCCTCACCA 173

Qy   1072 GCGGGGCTGATCAACGTCCACCCAGCAAGTCATCCACAGGCTTCTTTTAATCTGGC 1131
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   174 CAAATCCTGTGAATCCAGTGATCCTGAACAAAAGAAAACAGTAGCGTGGCCCTCCTCACCA 233

RESULT 15

```
LOCUS       BE376590               663 bp    mRNA    linear   EST 21-JUL-2000
DEFINITION  6012641911F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3584645 5',
            mRNA sequence.
ACCESSION   BE376590
VERSION     BE376590.1  GI:9321955
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 663)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM8743  row: f  column: 06
            High quality sequence stop: 498.
FEATURES             Location/Qualifiers
     source          1..663
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3584645"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT    143 a   193 c   170 g   157 t
ORIGIN
```

```
Query Match     15.4%;  Score 180.2;  DB 10;  Length 663;
Best Local Similarity   72.6%;  Pred. No. 6.6e-32;
Matches  233;  Conservative  0;  Mismatches  88;  Indels  0;  Gaps  0;

Qy   42 CCGGTTCTGGGGTCTGGGTAATGAGGTCATAGAAGGCCCCCAAAATGCAAAGAGTCCTGAAGG 101
        |||  ||||| ||| ||  |  || ||||| || |||| || || | | || ||| || ||||
Db   90 CAGCTTCCGGATCCAGTTTATCAGATCATTAGAGAGGTCCTCAGATGTAACAGTCCTRAAGG 149

Qy  102 GCTCCCAGGCTCGCTTCAAACTGCACCGTCTCCCAGGGCTCGAAGCTCATCATGTGGGCTC 161
        |||  |||||||| ||||||||| ||||||| | | ||| | |||||| |||||||| |||
Db  150 ACTCAGAGGCTCACTTCAAACTCCACCGTCGACTCACGCTGCAGAGCTTCTCATGTGACTC 209

Qy  162 TCAGTGACATGGTGGTGCTAAGGCGTCAGGCCCATGGAGCCCATGAATGACCGGCT 221
        | |  |||| ||||| ||| ||  ||| || ||||  | |||| |   || | ||
Db  210 TTAACCAAAATGTGGGTGCTGAGTCTCACCACCAAGGACCCATCATCAACAACGCT 269

Qy  222 TCACCTCTCAGAGGTACGGACCAGGGCGGGGACNTTCACCTCGGACGATGATCATNCCACAATG 281
        | ||| |||| ||| |  |  ||||   |||   ||||||| || | |||| || |  ||| |
Db  210 TTAACCAAAATGTGGGTGCTGAGTCTCACCACCAAGGACCCATCATCAACAACGCT 269
```

```
Db  295 TCACCTATGCCAGTTACAACAGCACACTGACACAGCGTTCATCTCGGAGTGATCATCCATGATG 354
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  282 TGGAGCCCAGTGATTCGGGGAACATCAGAGCCTTCCAGAAACAGTCGCCTTGCATGGAT 341
Db  355 TGCAGCCCAGTGACTCGGGATCCGTGCAATGCAGCCATGGGTTTCGGAT 414

Qy  342 CTGCTTACCTTACCGTCCCAAG 362
        |||| |||  || ||| |  ||
Db  415 CTGCCTTCCTCTCAGTGCCNAG 435
```

```
Db  270 TCACCTATGCCAGTTACAACAGCACACTGACACAGCGTTCATCTCGGAGTGATCATCCATGATG 329
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  282 TGGAGCCCAGTGATTCGGGGAACATCAGAGCCTTCCAGAAACAGTCGCCTTGCATGGAT 341
Db  330 TGCAGCCCAGTGACTCGGGATCCGTGCAATGCAGCCATGGGTTTCGGAT 389

Qy  342 CTGCTTACCTTACCGTCCCAAG 362
        |||| |||  || ||| |  ||
Db  390 CTGCCTTCCTCTCAGTGCAAG 410
```

Search completed: April 29, 2003, 09:24:15
Job time : 1902.29 secs

Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAAGAGGATCCCAAAAATGCAAGAGTCCTTAAAGG 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 5.0 and subtraction to
Rot = 20.0. Second strand cDNA was prepared with the primer adapter
of sequence[5'
GAGAGAGAAGAGGATCCTCGAGTTAATTAAAATTAATCCCCCCCCCC 3'], cDNA was cleaved
with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI.
Host: SOLR.

FEATURES   Location/Qualifiers
source   1..1552
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="FANTOM_DB:2010003D20"
                /db_xref="MGD:MGI:1904786"
                /clone="2010003D20"
                /db_xref="taxon:10090"
                /sex="male"
                /tissue_type="small intestine"
                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="adult"
CDS   73..885
                /note="data source:SPTR, source key:Q9NSI5, evidence:ISS
                homolog to IGSF5 PROTEIN (FRAGMENT)"
                /codon_start=1
                /protein_id="BAB25436.1"
                /db_xref="GI:12842010"
                /db_xref="MGD:MGI:1919308"
                /translation="MEGSWRDVLAVLVILAQLTRASGSSYQIIEGPQNVTVLKDSEAHF
                NCTVTHGWKLLMWLMWLLNQMVVLSLTTQGPIITNNRFTYASYNSTDSFISELIHDVQPS
                DSGSVQCSLQNSHGFGSAFLSVQDSIGEGPALPTWAILLAVAFSLILLIIVLIII
                FCCCASRRKEESTYQNEIRKSANMRTNKADPETKLKSGKENYGYSSDEARAAQTAS
                LPRKSAEVSLPEKRSSSLPYQELNKHQPGPATHPRVSFDIASPQKVRNVTLV"
polyA_signal   1534..1539
                /note="putative"
polyA_site   1552
                /note="putative"
BASE COUNT   430 a   369 c   374 g   379 t
ORIGIN

Query Match   15.6%;   Score 181.8;   DB 11;   Length 1552;
Best Local Similarity 72.9%;   Pred. No. 3.1e-32;
Matches 234;   Conservative 0;   Mismatches 87;   Indels 0;   Gaps 0;

Qy   42 CCGGTTCTGGGTCTCGGTAATGAAGTCATAGAAGGCCCCCAAAAATGCAAGAGTCCTTGAAGG 101
          ||||| |||||||||| |||||||||||||||||||| | ||| ||||||||||||| |||||
Db   128 CAGCTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAAATGTAACAGTCCTAAAGG 187

Qy   102 GCTCCCAGGCGTCGCTTCAACTGCACCGTCTCCCAGGCGTCGGAAGCTCATCATGTGGGGCTC 161
          |||||||||||||| ||||||||||| |||||||||| || |||||| |||||||||||||||
Db   188 ACTCAGAGGGTCACTTCAACTGCACCGTGACTCACGGGCTGGAAGCTTCTCATGTGGGACTC 247

Qy   162 TCAGTGACATGGTGGTGCTAAGCGTCAGGCCCATGGAGCCCCATCACCAACCAATGACCGCT 221
          |||||| |||||||||||||||||||||||||| ||||||||||||| |||||||||||||||
Db   248 TTAACCAAATGGTGGTGCTCAGTCTCCACCAAGGACCCCATCAACCATCACCAACAACCGCT 307

Qy   222 TCACCTCTCAGAGGTACGACCAGGCGGGGAACTTCACCTCCGGAGATGATCATCACACAATG 281
          |||||||| |||||||||||||||||| |||||||||||||| |||||||||||||||| |||
Db   308 TCACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATCCATGATG 367

Qy   282 TGGAGCCCAGTGATCGGGGAACATCAGATGCAGCCCTCCAGAAACAGAGTCGCCTGCATGGAT 341
          || ||||| || |||| |||||||| |||| ||||||| |||| |||||||| | |||||||
Db   368 TGCACCCCAGTGACTCGGGAATCCGTGCAATCGCCCCTGCGAGAGAACAGCCCATGGAGGGTTTGGGAT 427

Qy   342 CTGCTTACCTTACCGTCCCAAG 362
          |||| ||||||| | ||||||
Db   428 CTGCCTTCCTCTCAGTGCAAG 448

FEATURES   Location/Qualifiers
source   1..486
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="IMAGE:693713"
                /clone_lib="Soares mouse NML"
                /tissue_type="Liver"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                was primed with a Not I - oligo(dT) primer [5'
                TGTTACCAATTCTGAAGTGGGCAGGCGCGCCGGAACTCTTTTTTTTTTTTTTTTTTT 3'];
                double-stranded cDNA was ligated to Eco RI adaptors
                (Pharmacia), digested with Not I and cloned into the Not I
                and Eco RI sites of the modified pT7T3 vector. Library
                constructed and normalized by Bento Soares and M.Fatima
                Bonaldo."
                /lab_host="DH10B"
BASE COUNT   119 a   133 c   126 g   107 t   1 others
ORIGIN

Query Match   15.5%;   Score 180.8;   DB 9;   Length 486;
Best Local Similarity 72.6%;   Pred. No. 4.6e-32;
Matches 233;   Conservative 0;   Mismatches 88;   Indels 0;   Gaps 0;

Qy   42 CCGGTTCTGGGTCTCGGTAATGAAGTCATAGAAGGCCCCCAAAAATGCAAGAGTCCTTGAAGG 101
          ||||| |||||||||| |||||||||||||||||||| | ||| ||||||||||||| |||||
Db   115 CAGCTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAAATGTAACAGTCCTAAAGG 174

Qy   102 GCTCCCAGGCGTCGCTTCAACTGCACCGTCTCCCAGGCGTCGGAAGCTCATCATGTGGGGCTC 161
          |||||||||||||| ||||||||||| |||||||||| || |||||| |||||||||||||||
Db   175 ACTCAGAGGGTCACTTCAACTGCACCGTGACTCACGGGCTGGAAGCTTCTCATGTGGGACTC 234

Qy   162 TCAGTGACATGGTGGTGCTAAGCGTCAGGCCCATGGAGCCCCATCACCAACCAATGACCGCT 221
          |||||| |||||||||||||||||||||||||| ||||||||||||| |||||||||||||||
Db   235 TTAACCAAATGGTGGTGCTCAGTCTCCACCAAGGACCCCATCAACCATCACCAACAACCGCT 294

Qy   222 TCACCTCTCAGAGGTACGACCAGGCGGGGAACTTCACCTCCGGAGATGATCATCACACAATG 281
          |||||||| |||||||||||||||||| |||||||||||||| ||| |||| || |||| |||
Db   281 ...

```
                    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                    DNA Sequencing by: Incyte Genomics, Inc.
                    Clone distribution information can be
                    found through the I.M.A.G.E. Consortium/LLNL at:
                    http://image.llnl.gov
                    Plate: LLAM10260  row: n  column: 06
                    High quality sequence stop: 628.
FEATURES             Location/Qualifiers
     source          1..931
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4455973"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor; biopsy sample"
                     /dev_stage="10 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                     library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT    233 a    277 c    242 g    179 t
ORIGIN

Qy      42 CCGGTTCTTCGGGTCTTGGTTTAATGAAGTCATAGAAGAGGCCCCAAAATGCAAGAGTCCTGAAGG 101
            | || ||||  |||||||||||||| |||||| ||||||||||||||||||||| ||||||||||
Db     156 CAGGTTCTTCGGATCCAGTTATTCAGATCATAGAAGAGTCCTTCAGAAATGTAACAGTCCTAAAGG 215

Qy     102 GCTCCCAGGCCTCGCTTCAACTGCACCGTCTCCCAGGGGTCTGGGAAGCTCATCATGTGGGCCTC 161
            ||||| ||||||||| |||||| |||| |||| ||  |||| | |||| ||||||||| ||||||
Db     216 ACTCAGAGGCTCACTTCAACCTGCACCGTCGACCTGGAAGCTTCTCATGTGTGGACTC 275

Qy     162 TCAGTGACATGGTGGTTGCTAAGCGTCAGGCCCCATGGAGCCCCATCATCACCAATGACCGGCT 221
            | ||||||||||||||||| |||||||||||| || ||| ||||||||| |||| |||||||||
Db     276 TTAACCAAATGGTGGTCTGTGAGTCTCACCACCAAGGACCATCATCACCAACAACCGCT 335

Qy     222 TCACCTCTCAGAGGTACGACCCAGGGGCGGGAACTTCACCTCGGAGATGAATCATCCACAATG 281
            |||||||||||| |||| ||||| ||||| |||||||| |||||||||||| ||||||||||
Db     336 TCACCTATGCCAGTTACAACAGCACTTCAATCTCTCGGAGTTGAATCATCCATCATG 395

Qy     282 TGGAGCCCCAGTGATTCGGGGAACATCAGATGCAGCCCTCCAGAACAGTCGCCTGCATGGAT 341
            | |||||| ||||  ||||||||||| |||| ||| |||||||| ||||| || |||| |||
Db     396 TGCAGCCCCAGTGACTCGGGGATCCGTGCAAATGCAGCCACCAGGAACAGCCCATGGGTTGGGAT 455

Qy     342 CTGCTTACCTTACCGTCCAAG 362
            |||| | |||| ||| ||||
Db     456 CTGCCTTCCTTCCAGTGCAAG 476


RESULT 13
LOCUS       AK008060                1552 bp    mRNA    linear   HTC 19-JAN-2002
DEFINITION  Mus musculus adult male small intestine cDNA, RIKEN full-length
            enriched library, clone:2010003D20:homolog to IGSF5 PROTEIN
            (FRAGMENT), full insert sequence.
ACCESSION   AK008060
VERSION     AK008060.1  GI:12842009
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

```
REFERENCE   1
  AUTHORS   Carninci,P. and Hayashizaki,Y.
  TITLE     High-efficiency full-length cDNA cloning
  JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE   99279253
```

```
REFERENCE   2
  AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE   20499374
   PUBMED   11042159
REFERENCE   3
  AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
  JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE   20530913
   PUBMED   11076861
REFERENCE   4
  AUTHORS   Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J.,
            Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C.A., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
  TITLE     Functional annotation of a full-length mouse cDNA collection
  JOURNAL   Nature 409 (6821), 685-690 (2001)
  MEDLINE   21085660
   PUBMED   11217851
REFERENCE   5  (bases 1 to 1552)
  AUTHORS   Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
            Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
            Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
            Hara,A., Hayatsu,N., Hill,D., Hiramoto,M., Hiraoka,H., Hori,F.,
            Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
            Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
            Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
            Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
            Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
            Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
            Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I.,
            Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
```

Left column (RESULT 11)

FEATURES
  source

MGI:427273
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 454.

    1. .471
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="IMAGE:693713"
    /clone_lib="Soares mouse NML"
    /tissue_type="Liver"
    /lab_host="DH10B"
    /note="vector: pT7T3D-Pac (Pharmacia) with a modified
    polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
    was primed with a Not I - oligo(dT) primer [5'
    TGTTACCAATCTGAAGTGGGAGGCGGCCGCAGAAATCTTTTTTTTTTTTTTTT 3'];
    double-stranded cDNA was ligated to Eco RI adaptors
    (Pharmacia), digested with Not I and cloned into the Not I
    and Eco RI sites of the modified pT7T3 vector. Library
    constructed and normalized by Bento Soares and M.Fatima
    Bonaldo."

BASE COUNT   118 a   126 c   122 g   105 t
ORIGIN

Query Match   Best Local Similarity 72.9%; Score 181.8; DB 9; Length 471;
Matches 234; Conservative 0; Mismatches 87; Indels 0; Gaps 0; Pred. No. 2.7e-32;

```
Qy    42 CCGGTTCTCGGGTCTGGTAATGAAGTCATAGAAGGCCCCCAAAAATGCAAGAGTCTGAAGG  101
Db   115 CAGCTTCCCGGATCCAGTTATCAGATCATAGAAAGGTCCTCAGAAGTAACAGTCCTAAAGG  174
Qy   102 GCTCCCAGGCGTCGCGTTCAACTGCACCGTCTCCCAGGGGTCGGGAAGCTCATCATGTGGGGCTC  161
Db   175 ACTCAGAGGCGTCACTTCAACTGCACCGTCGACTCACGGCTGGAAGCTTCTCATGTGGACTC  234
Qy   162 TCAGTGACAATGGTGGTGCTTAAGCGTGTCAGGCCCATGGAGCCCATCATCACCAATGACCGGCT  221
Db   235 TTAACCAAAATGGTGGTGCTCGAGTCTCCACCAGGACCCATCATCACCAACAACCGCT  294
Qy   222 TCACCCTCTCAGAGGTACGACCAGGCGCGGGAAACTTCACCCTCGGGAGATGACATCACACAATG  281
Db   295 TCACCTATGCCAGTTACAACAGCACACTGACAGCTTCATCTCGGAGTTGATCATCCATGATG  354
Qy   282 TGGGAGCCCAGTGATTCGGGGGAACATCAGATGCAGCAGTCGCCTGCATGGAT  341
Db   355 TGCAGCCCAGTGACTCGGGATCCGGTGCAATGCAGCCTGCAGAGAACAGCCATGGGTTTGGAT  414
Qy   342 CTGCTTACCTTACCGTCCAAG  362
Db   415 CTGCCTTCCTCTCTCAGTGCAAG  435
```

RESULT 11
LOCUS   AW990468   611 bp   mRNA   linear   EST 02-JUN-2000
DEFINITION uf37d05.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
    IMAGE:1513545 5' similar to TR:O60962 O60962 L549.1. ;, mRNA
    sequence.
ACCESSION AW990468
VERSION   AW990468.1 GI:8186027
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 611)
  AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
    Tumor Gene Index
  JOURNAL Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-r@mail.nih.gov
    This clone is available royalty-free through LLNL ; contact the

Right column (RESULT 12)

FEATURES
  source

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:940397
Seq primer: -40RP from Gibco
High quality sequence stop: 474.

    1. .611
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="IMAGE:1513545"
    /clone_lib="Soares_mammary_gland_NMLMG"
    /sex="female (lactating)"
    /tissue_type="mammary gland"
    /lab_host="DH10B"
    /note="vector: pT7T3D-Pac (Pharmacia) with a modified
    polylinker; 1st strand cDNA was prepared from mammary
    gland tissue from a lactating female, and was then primed
    with a Not I - oligo(dT) primer. Double-stranded cDNA was
    ligated to Eco RI adaptors (Pharmacia), digested with Not
    I and cloned into the Not I and Eco RI sites of the
    modified pT7T3 vector. Library is normalized. Library
    was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT   152 a   155 c   156 g   148 t
ORIGIN

Query Match   Best Local Similarity 72.9%; Score 181.8; DB 10; Length 611;
Matches 234; Conservative 0; Mismatches 87; Indels 0; Gaps 0; Pred. No. 2.8e-32;

```
Qy    42 CCGGTTCTCGGGTCTGGTAATGAAGTCATAGAAGGCCCCCAAAAATGCAAGAGTCTGAAGG  101
Db   124 CAGCTTCCCGGATCCAGTTATCAGATCATAGAAAGGTCCTCAGAAGTAACAGTCCTAAAGG  183
Qy   102 GCTCCCAGGCGTCGCGTTCAACTGCACCGTCTCCCAGGGGTCGGGAAGCTCATCATGTGGGGCTC  161
Db   184 ACTCAGAGGCGTCACTTCAACTGCACCGTCGACTCACGGCTGGAAGCTTCTCATGTGGACTC  243
Qy   162 TCAGTGACAATGGTGGTGCTTAAGCGTGTCAGGCCCATGGAGCCCATCATCACCAATGACCGGCT  221
Db   244 TTAACCAAAATGGTGGTGCTCGAGTCTCCACCAGGACCCATCATCACCAACAACCGCT  303
Qy   222 TCACCCTCTCAGAGGTACGACCAGGCGCGGGAAACTTCACCCTCGGGAGATGACATCACACAATG  281
Db   304 TCACCTATGCCAGTTACAACAGCACACTGACAGCTTCATCTCGGAGTTGATCATCCATGATG  363
Qy   282 TGGGAGCCCAGTGATTCGGGGGAACATCAGATGCAGCAGTCGCCTGCATGGAT  341
Db   364 TGCAGCCCAGTGACTCGGGATCCGGTGCAATGCAGCCTGCAGAGAACAGCCATGGGTTTGGAT  423
Qy   342 CTGCTTACCTTACCGTCCAAG  362
Db   424 CTGCCTTCCTCTCTCAGTGCAAG  444
```

RESULT 12
LOCUS   BG173684
LOCUS   BG173684   931 bp   mRNA   linear   EST 06-FEB-2001
DEFINITION 602336823F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4459973 5',
    mRNA sequence.
ACCESSION BG173684
VERSION   BG173684.1 GI:12680387
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 931)
  AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-r@mail.nih.gov
    Tissue Procurement: Gilbert Smith, Ph.D.
    cDNA Library Preparation: Life Technologies, Inc.

COMMENT   1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
           Tel:81-45-503-9111, Fax:81-45-503-9170)
           Clones are derived from the chimpanzee PTB This BAC end
           was generated during the R&D process and may have higher chance of
           clone tracking errors.

FEATURES   Location/Qualifiers
    source   1..677
                   /organism="Pan troglodytes"
                   /db_xref="taxon:9598"
                   /clone="PTB-112N07.F"
                   /sex="male"
                   /cell_type="lymphoblast"
                   /clone_lib="PTB Chimpanzee Male BAC Library"

    PRIMERS
           Sequencing: -21M13
    LIBRARY
           Vector   : pKS145
           R.Site 1 : SacI
           R.Site 2 : SacI.

BASE COUNT   168 a   159 c   143 g   207 t
ORIGIN

Query Match   16.3%;   Score 190.0; DB 17; Length 677;
Best Local Similarity   94.3%;   Pred. No. 2.1e-34;
Matches 198; Conservative   0; Mismatches 12; Indels   0; Gaps   0;

Qy   467 TCCTGGGAGCTCGGTCTCCCTGGTCAGCCATTCAAGCTATTATTATTTGTTCCGGAGCCCAGC 526
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   72 TCCGCGGAGCTCGGTCTCCTGTCAGCCATTCAAGCTATTATTATTTGTTCCGGAGCCCAGC 131

Qy   527 GACCTTCAAAGTGCAGTGAGCAATCCTGGCTCTGACCCCCACAGAGCAATGGGACTTTGACT 586
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   132 GACCTTCAAAGTGCAGTGAGCAATCCTGGCTCTGACCCCCACAGAGCAATGGGACTTTGACT 191

Qy   587 TGCGTGGCTACCTGTGGAGAAGAGCCTGAAGGACCCCCACCAAGTGTCGAACTCTGTAAAATCTCACTGTG 646
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   192 TGCGTGGCTACCTGTGGAGAAGAGCCTGAAGGACCCCCACCAAGTGTCGAACTGTAAAATCTCACTGTG 251

Qy   647 ATTCGGTGTCTCCCCAAGACACTGGAGGTGGT 676
          ||||||||||||||||||||||||||||
Db   252 ATTCGGGCTCCCCCAAGGTAAGTGAAGATGT 281

REFERENCE   1 (bases 1 to 524)
   AUTHORS   Takasuga,A., Hirotsune,S., Itoh,R., Jitohzono,A., Suzuki,H., Aso.H.
             and Sugimoto,Y.
   TITLE   Establishment of a high throughput EST sequencing system using
             poly(A) tail-removed cDNA libraries and determination of 36,000
             bovine ESTs
   JOURNAL   Nucleic Acids Res. 29 (22), E108 (2001)
   MEDLINE   21570554
   COMMENT   Contact: Yoshikazu Sugimoto
             Animal Genetics Division
             Shirakawa Institute of Animal Genetics
             Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
             Tel: 81-248-25-5641
             Fax: 81-248-25-5725
             Email: kazusugi@cocoa.ocn.ne.jp
             Single pass sequencing.

RESULT 9
LOCUS   AV601192   524 bp   mRNA   linear   EST 27-NOV-2001
DEFINITION AV601192 Bos taurus kidney fetus Bos taurus cDNA clone E1KI002H04
            5', mRNA sequence.
ACCESSION   AV601192
VERSION   AV601192.1 GI:9723510
KEYWORDS   EST.
SOURCE   cow.
  ORGANISM Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.

FEATURES   Location/Qualifiers
    source   1..524
                   /organism="Bos taurus"
                   /db_xref="taxon:9913"
                   /clone="E1KI002H04"
                   /clone_lib="Bos taurus kidney fetus"
                   /tissue_type="kidney"
                   /dev_stage="fetus"
                   /lab_host="DH10B"
                   /note="Vector: pZL1; Site_1: SalI; Site_2: NotI; Poly A
                   was deleted from a NotI site"

This clone was obtained from a polyA-deleted cDNA library.

BASE COUNT   120 a   135 c   168 g   96 t   5 others
ORIGIN

Query Match   15.9%;   Score 185.8; DB 10; Length 524;
Best Local Similarity   75.3%;   Pred. No. 3.1e-33;
Matches 229; Conservative   0; Mismatches 75; Indels   0; Gaps   0;

Qy   42 CCGGTTCTTGGGTCTTGGTTAATGAAGTCATTAGAAGAGGCCCCCAAAAATGCAAGAGTCCTGAAGG 101
          ||| | ||| |||| | | | || || | || | ||| |||| | | ||||| |||||||||||
Db   221 CAGCCCTTTGGGATCCACCAGTGAAAATCATTAGAGAGGGTCCCAAGAAATGTCACAGCCCTGAAGG 280

Qy   102 GCTCCCAGGCGTCGGTCTCAACTGCACCGTCTCCCCAGGGCTGGAAGGTCTCATGTTGGGCTC 161
          | | | ||| ||| | | ||||| |||| |||||||||| ||||||||| |||| ||| ||
Db   281 GCTCGGAGGCTCGGCTTCAACTGCACCATCTCCCAGGGCTGGAAGGTCCTGTCATGTTGGGCTC 340

Qy   162 TCAGTGACATGGTGGTGGTTGCTAAGCGTCAGGCCCCATGGAGCGCCCATCATCACCAATGACCGCT 221
          ||| |||| | ||| | ||| | ||| | ||||| | ||| | || | || || ||||| ||||||
Db   341 TGAGAGGCCACAGTGGGTGCTGAGCATGACACCTAAATGAGACCCATCATCACCAGTGACCGCT 400

Qy   222 TCACCTCTCAGAGGTACGGACCAGGGCCGCGGGAACTTCACCTCGGAGGATGATGATCCACAATG 281
          ||| | | ||| ||| |||||||| | | | ||||| || || | ||| | || | || |||
Db   401 TCACTTCGGCAAGCTACCACCAAGAGGGCGGGGAACTTCATCTNTGAGATGATAATTCATGACG 460

Qy   282 TGGAGCCCAGTTGATTTGGGGAACATCAGATGCAGACGATCCCAGAAACAGTCGCGCCTGCATGGAT 341
          | | | ||| || | ||||||| ||||||||| | || | |||| | || ||| | || | || ||
Db   461 TGCAACTCAGCCGATGCGGGCGCAAGTNAAATGCAGCCTTCANAACAGCAATGGGGATGGAG 520

Qy   342 CTGC 345
          | | |
Db   521 CCGC 524

REFERENCE   1 (bases 1 to 471)
   AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R.
   TITLE   The Washu-HHMI Mouse EST Project
   JOURNAL   Unpublished (1996)
   COMMENT   Contact: Marra M/Mouse EST Project
             Washu-HHMI Mouse EST Project
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.

RESULT 10
LOCUS   AA265274   471 bp   mRNA   linear   EST 20-MAR-1997
DEFINITION mx91609.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:693713
            5', mRNA sequence.
ACCESSION   AA265274
VERSION   AA265274.1 GI:1901370
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

USDA, ARS, Beltsville Agricultural Research Center
Bldg. 200 Rm 2A, Beltsville, MD 20705, USA
Tel: 301 504 8416
Fax: 301 504 8414
Email: tads@anri.barc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAAACAGCTATGACCAT
BACKWARD: GTTTTCCCAGTCACGACG
Plate: 27 row: H column: 19
Seq primer: ATTTAGGTGACACTATAG.

```
FEATURES
          source          1..517
                          /organism="Bos taurus"
                          /db_xref="taxon:9913"
                          /clone_lib="BARC 5BOV"
                          /tissue_type="pooled"
                          /lab_host="DH10B"
                          /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                          Library made from pooled mRNA isolated from mammary
                          tissues at eight physiological, developmental, and disease
                          states."
BASE COUNT     117 a    133 c    165 g    102 t
ORIGIN

Query Match    16.5%;   Score 193.2;  DB 10;   Length 517;
Best Local Similarity   75.5%;  Pred. No. 5.6e-35;
Matches 240;  Conservative 0;  Mismatches 78;  Indels 0;  Gaps 0;

Qy     42 CCGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGCCCCAAAAATGCAAGAGTCCTTGAAGG 101
          || ||| ||||||| |||| || |||||||| || ||||| || ||| |||||||  |||||
Db    200 CAGCCTGTGGATCCAGCAGTGAAATCATAGAGGGTCCCAAGAAATGTCACAGCCCTTGAAGG 259

Qy    102 GCTCCCAGGCGTCGTCGTTCAACTGCACCGTGCTCTCCCAGGGCTGGGAAGCTCAATCGTGGGGCTC 161
          ||||| |||||||||  ||||||||||||| | |||||||||||||  |||||| || ||||||||||
Db    260 GCTCCGGAGGCGTCGCTTCAACTGCACCATCTCCCAGGGCTGGAAGCTCGTCTCATGTGGGGCTC 319

Qy    162 TCAGTGACATGGTGGTGCTAAGCCCCATGGAGCCCATCATCACCAATGACCGGT 221
          |||||||||| |||||||||||| || ||||| |||||||||||||||||||||
Db    320 TGAGAGGCACAGTGGTGCTGAGCATGACCATCATCATCACCAGTGACCGCT 379

Qy    222 TCACCTCTCAGAGGTACGACCACGGGCGGGGGAACTTCACCTCGGGAGATGATCATCCACAATG 281
          |||||||||||||| || |||||||||||||||||||||||||| | ||||| || |||||||
Db    380 TCACTTCGGCAAGCTACCAAGAGGGGGGAACTTCATCTCTGAGATGATAAATTCATGACG 439

Qy    282 TGGGAGCCCCAGTGATTCGGTGGGAACATCAGATGACGATGCAGCCTCTCGAGACAGTCGCCTGCATG 341
          ||||||||||||||||||||| ||||||| ||||| |||||||||||||||||||||||||||||||
Db    440 TGCAACTGAGCGAATGCCGGGCAAGTCAAATGCAAGCCTTCAGAAGCAGAATCGGGATGGGAG 499

Qy    342 CTGCTTACCTTACCGTCC 359
          ||||| ||||||||||||
Db    500 CCGCCTTCCTTTCTGTC 517
```

REFERENCE   2 (bases 1 to 622)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:chimpbes@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)

COMMENT   Clones are derived from the chimpanzee BAC library PTB This BAC end
            was generated during the R&D process and may have higher chance of
            clone tracking errors.
            PRIMERS
            Sequencing: 21M13

```
FEATURES
          source          1..622
                          /organism="Pan troglodytes"
                          /db_xref="taxon:9598"
                          /clone="PTB-060F12.F"
                          /sex="male"
                          /cell_type="lymphoblast"
                          /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT     165 a    153 c    128 g    176 t
ORIGIN

Query Match    16.3%;   Score 190.8;  DB 17;   Length 622;
Best Local Similarity   94.3%;  Pred. No. 2.1e-34;
Matches 198;  Conservative 0;  Mismatches 12;  Indels 0;  Gaps 0;

Qy    467 TCCTGGGAGCTCGGTCTCCTGGGTCAGCCAATTCAAGCTATTATTTTGTGTTCCGGAGCCCAGC 526
          ||| ||||||||||||||||||| |||||| |||||||| |||||||| ||||||||||||||||
Db     72 TCCCGGGAGCTCGGTCTCCTGGTCAGCCATTCAAGCCATTATTTTGTTCCGGAGCCCAGC 131

Qy    527 GACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGAGCCCACAGAGCAAATGGGACTTTGACT 586
          ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    132 GACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGACCCCACAGAGCAAATGGGACTTTGACT 191

Qy    587 TGCCTGGGCTACCTGGAAGAGAGCCTGAAAGGCCCGCAAAGTCTGCAACTGTGAAATCTCACTGTG 646
          |||||||||||||||||||||||||||||||||| |||||||||||||||| |||||||||||||
Db    192 TGCCTGGGCTACCTGGAAGAGAGCCTGAAAGGCCCACAAAGTCTGCAACTGTAAATCTCACTGTG 251

Qy    647 ATTCGGTGTCCCCAAGACACTGGAGGTGGT 676
          ||||||  ||||||||| |  || |||| ||
Db    252 ATTCGGCGTCCCCAAGGTAAGTGAAGATGT 281
```

RESULT 7
```
LOCUS       AG069679       622 bp    DNA    linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-060F12.F, genomic survey sequence.
ACCESSION   AG069679
VERSION     AG069679.1  GI:16621481
KEYWORDS    GSS.
SOURCE      Pan troglodytes
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
```
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   BAC end sequences of Library PTB
  JOURNAL   Unpublished

RESULT 8
```
LOCUS       AG107877       677 bp    DNA    linear   GSS 03-NOV-2001
DEFINITION  Pan troglodytes DNA, clone: PTB-112N07.F, genomic survey sequence.
ACCESSION   AG107877
VERSION     AG107877.1  GI:16728395
KEYWORDS    GSS.
SOURCE      Pan troglodytes
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
```
REFERENCE   1
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   BAC end sequences of Library PTB
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 677)
  AUTHORS   Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
            Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE   Direct Submission
  JOURNAL   Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);

(Column 1)

primed with a primer [5'
GAGAGAGAGGACGGGCGCCCAACTCGAGTTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGATTCTCGAGTTAATTAAATTAAATTCCCCCCCCCCC 3']. cDNA
was cleaved with BamHI and XhoI. Vector: a modified
pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT   154 a   167 c   161 g   149 t
ORIGIN

Query Match   22.0%;   Score 256.8;   DB 10;   Length 631;
Best Local Similarity   71.8%;   Pred. No. 6.3e-50;
Matches 336;   Conservative 0;   Mismatches 132;   Indels 0;   Gaps 0;

```
Qy  42 CCGGTTCTTGGGTCTGGTTAATGAAGTCATAGAAGAGAGGGCCCCAAAAATGCAAGAGTCCTGAAGG 101
Db 161 CAGCTTCCGGATCCAGTTATCAGATCAATAGAAGAGGTCCTCAGAAGAATGTAACAGTCCTAAAGG 220

Qy 102 GCTCCCAGGCTCGCTTCAAACTCGACACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCTC 161
Db 221 ACTCAGAGGCTCACTCAACTCGACACCGTCTGGTAGGCTCATCATGTGACTC 280

Qy 162 TCAGTGACATGGTGGTGCTAAGCGTCAGGCCCATGGAGCCCAATCATCACCAAATGACCGGCT 221
Db 281 TTAAACAAATGGTGAGTGGGTGCTGAGTGTCTCACCACCAAGGACCACCAAACCACCGCT 340

Qy 222 TCACCTCTCAGAGGTACGACCAGGCGGGAACTTCACCTCGGAGATGATCATCACAATG 281
Db 341 TCACCTATGCCAGTTAACAACAGCACAGTTCATCTCGGAGTTGATCATCCATGATG 400

Qy 282 TGGAGCCCAGTGATTCGGGGAACATCAGATGACAGTTCATCCATCATCCATGATG 341
Db 401 TGCAGCCCAGTGACTCGGTGCAAATGCAGCCTGCAAATGCAGCCGTCAAATGCTTCCATGATG 460

Qy 342 CTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTTCATTCCCAGTGTTAATCTTGTAG 401
Db 461 CTGCCTTCCTCTTCACGTCGCAAGTCATGGGGACCCTGAACATTCCTAGCAACAACCTTATAG 520

Qy 402 TCGCTGAGAATGAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGACCTGGCTCCCGG 461
Db 521 TCACTGAGGTGAACCCTGTAATGTGACTTGTTATGCCGTGGGCTGGGCTCCACTCCCCG 580

Qy 462 ATATTTCCTGGGAGCTCGGGTCTCCTCTGTCTGTCCAGCCATTCAAGCTATTATT 509
Db 581 ATATTTCCTGGGAGCTTGAGGTTCCGTAAGCCATTCGAGTTACAAT 628
```

.

RESULT 5
BE032610
LOCUS   BE032610   527 bp   mRNA   linear   EST 09-JUL-2000
DEFINITION 132035 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BE032610
VERSION   BE032610.1   GI:8327619
KEYWORDS   EST.
SOURCE   pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1 (bases 1 to 527)
AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
TITLE   Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
JOURNAL   Unpublished (2000)
COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov

(Column 2)

Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
  FORWARD: AGGAAACAGCTATGACCAT
  BACKWARD: GTTTTCCCAGTCACGACG

FEATURES   Location/Qualifiers
     source   1..527
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."
BASE COUNT   105 a   147 c   176 g   99 t
ORIGIN

Query Match   20.7%;   Score 242;   DB 10;   Length 527;
Best Local Similarity   78.4%;   Pred. No. 1.9e-46;
Matches 290;   Conservative 0;   Mismatches 80;   Indels 0;   Gaps 0;

```
Qy  42 CCGGTTCTTGGGTCTGGTTAATGAAGTCATAGAAGAGAGGGCCCCAAAAATGCAAGAGTCTGAAGG 101
Db 158 CAGCTTGTGGGATCCGGCAATGAAAATCTTAGGAGGGTCCAGGAGAACGTCACGGCGGTTGAAGG 217

Qy 102 GCTCCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCTC 161
Db 218 GCTCGGAGGCTCGGCTTCAAACTGCACCATCTCCAGGGCTGGAAGCTCATCATGTGGGCTC 277

Qy 162 TCAGTGACATGGTGGTGCTAAGCGTCAGGCCCATGGAGCCCAATCATCACCAAATGACCGGCT 221
Db 278 TGAAAGGCACGGTGGTTGCGTCACGCCCACCAGCCCCATCATCACCACAAACGACCGAT 337

Qy 222 TCACCTCTCAGAGGTACGACCAGGCGGGAACTTCACCTCGGAGATGATCATCACAATG 281
Db 338 TCACGTCTGCAAGTTATGAAAGAGGGCGTAACTTCATCTCTGAGATGATCATCCATGACG 397

Qy 282 TGGAGCCCAGTGATTCGGGGAACATCAGATGACAGTTCATCCATCATCCATGATG 341
Db 398 TGCAGCTCAGTGATGCCGGGCAGGTCAAAATGCAGCCTCCAGAACACTGACGGGATGGAT 457

Qy 342 CTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATCTTGTAG 401
Db 458 CCGCTTTCCTTTCCGTCCAAGTCATGGGGTGTCCTCATTCCCGGAGCAGCCGCTTCGTAG 517

Qy 402 TCGCTGAGAA 411
Db 518 TCCTTGCGGA 527
```

RESULT 6
BE476432
LOCUS   BE476432   517 bp   mRNA   linear   EST 28-AUG-2000
DEFINITION 159670 BARC 5BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   BE476432
VERSION   BE476432.1   GI:9595965
KEYWORDS   EST.
SOURCE   cow.
ORGANISM   Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
REFERENCE   1 (bases 1 to 517)
AUTHORS   Sonstegard,T.S., Capuco,A.V., Van Tassell,C.P., Ashwell,M.S. and
            Wells,K.D.
TITLE   Mapping of Expressed Sequence Tags from a normalized bovine mammary
            gland cDNA library
JOURNAL   Unpublished (2000)
COMMENT   Contact: Sonstegard TS

```
Db    628  C--GGGCACTTTATGAGGGTCTGGAGGTCGAGTGTCCTGGAACTCA-CACACTGGGCACGGGACT
                |||||        ||||| || ||| || |  ||| | ||  |       |||| |||||||||| |
Qy    582  TGACTTGCTGCTTACCTGGAAGAGCCTGAAGGCCCGCGAAGTCT 625
                |||||| |  || ||||| ||  ||| || |  ||| | ||||
Db    685  TGACTTGTGTGGCAAGAGCTGAAGAGCTTGCAGGCCAGCAGTCT 728
```

RESULT 3

LOCUS   BG740428   725 bp   mRNA   linear   EST 15-MAY-2001
DEFINITION 602633817F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4778789 5',
            mRNA sequence.
ACCESSION   BG740428
VERSION   BG740428.1 GI:14051081
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 725)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/;
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10635 row: b column: 06
            High quality sequence stop:725.
FEATURES   Location/Qualifiers
     source   1..725
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4778789"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   227 a   163 c   159 g   176 t
ORIGIN

Query Match   22.9%; Score 267.4; DB 12; Length 725;
Best Local Similarity 99.6%; Pred. No.2e-52;
Matches 268; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   899 AAGAAAATCTGAAAAAGAGAAGACAAACAAAAGAAAACTGAGACAGAAAGTGGAAATGAAAAC 958
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   377 AGGAAAATCTGAAAAAGAGAAGACAAACAAAAGAAAACTGAGACAGAAAGTGGAAATGAAAAC 436

Qy   959 TCCGGCTACAAATTCAGATGAACAAAAGACCACACAACACCGGCGTTCTCTCCCCTCCCAAATCC 1018
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   437 TCCGGCTACAAATTCAGATGAACAAAAGACCACACAACACCGGCGTTCTCTCCCCTCCCAAATCC 496

Qy   1019 TGTGAATCCAGTGATCCTGAACTCGTGGCCCTCCTCCACGCGGGGCT 1078
                ||||||||||||||||||||||||||||||||||||||||||||||
Db   497 TGTGAATCCAGTGATCCTGAACAAAAGAAACAGTAGCTGTGGCCCTCCTCCACGCGGGGCT 556

Qy   1079 GATCAAACGTCCACCCAGGCCAGCAAGTCATCCACAGGCGTTCTT 1138
                |||||||||||||||||||||||||||||||||||||||||||||
Db   557 GATCAAACGTCCACCCAGGCCAGCAAGTCATCCACAGGCGTTCTT 616

Qy   1139 GAGAAGGTCAGTAATACAACTGTAGTATA 1167
                ||||||||||||||||||||||||||||
Db   617 GAGAAGGTCAGTAATACAACTGTAGTATA 645
```

RESULT 4

LOCUS       BB663870                 631 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB663870 RIKEN full-length enriched, 0 day neonate lung Mus
            musculus cDNA clone E030016M12 5', mRNA sequence.
ACCESSION   BB663870
VERSION     BB663870.1  GI:16497624
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 631)
AUTHORS     Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL     Unpublished (2001)
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC) Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp;
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res., 10 (10), 1617-1630 (2000)
            Watahiki,M., Yoneda,Y., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            wagi,K., Fujiwake,S., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res., 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES             Location/Qualifiers
     source          1..631
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="E030016M12"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     lung"
                     /tissue_type="lung"
                     /dev_stage=0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
```

source
1. .784
/organism="Mus musculus"
/strain="NMRI"
/db_xref="taxon:10090"
/clone="IMAGE:4917803"
/clone_lib="NCI_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Primer: Oligo dT.
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Site_1: SalI;
Library constructed by Life Technologies. Investigators
providing samples: Lothar Hennighausen/Priscilla Furth,
NIH Reference for transgenic model: Li et al., Cell Growth
and Differentiation 7,3-11 (1996)."

BASE COUNT   188 a   202 c   208 g   186 t
ORIGIN

Query Match   26.2%;   Score 306; DB 12;   Length 784;
Best Local Similarity   70.6%;   Pred. No. 1.7e-61;
Matches 436;   Conservative   0;   Mismatches 180;   Indels   2;   Gaps   2;

```
Qy     42 CCGGTTCTGGGTCTGGTAATGAAGTCATAGAAGAGGCCCCAAAATGCAAGAGTCCTGAAGG 101
          ||||| || ||||| |||  |||  ||| ||| || |||||| |||| || |||||||||||
Db    148 CAGCTTCCGGATCCAGTTATCAGATCATAGAAGTGTAACAGTCCTAAAGG 207
Qy    102 GCTCCCAGGCCTGCTTCAACTGCACCGTCTCCCAGGCCTGGAAGCTCATCATGTGGGCTC 161
Db    208 ACTCAGAGGCTCACTTCAACTGCACCGTGACTCACGGCTGGAAGCTTCTCATGTGGACTC 267
Qy    162 TCAGTGACATGGTGGTGGGTTCAGGCCATGGGAGACCCCATCATCACCAATGACCGGCT 221
Db    268 TTAACCAAAATGGTGGTGCTGAGTCTCACCACCAAGACCATCATCACCAACAACCGCT 327
Qy    222 TCACCTCTCAGAGGTACGACACCAGGCGGGGAACTTCACCTCGGAGATGATCATCACAATG 281
Db    328 TCACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATCCAATCATGAATG 387
Qy    282 TGGAGCCCCAGTGAATTCGGGGACATCAGATGCAGCCTTCCAGAACAGCAGTGTCCTGCATGGAT 341
Db    388 TGCAGCCCCAGTGACTCGGGATCCGTGCAATCAGCCCACAGCCGGTTTGGAT 447
Qy    342 CTGCTTACCTTACGGTCGCAAGTGTTATTGGGGAGAGGCGTGTTCATTCCCAGTGTTAATCTTGTAG 401
Db    448 CTGCCTTCCTCTCAGTGCAAGTCATGGGGACCCTGACACTTCCTAGCAACAACCTTATAG 507
Qy    402 TCGCTGAGAATGAAACCTTGTGAAGTTACTTGTTGCTACCCTCACACTGGACCTGGCTCCCGG 461
Db    508 TCACTGAGGGTGAACCCTGTAAATGTGACTTGAACCTCACTCCCGG 567
Qy    462 ATATTTCCTGGGGACTGGGGTGTCCTCCTGGTCAGCCATTCAAGTCTATTATTTTGTTCCGAGC 521
Db    568 ATATTTCCTGGGAGCTTGAGGTTCCCGTTACAAATTCCTTTCTGGAGC 627
Qy    522 CCAGCGACCTTCAAAGTCAGCGAGCATCAGAGAGCAATGGGACTT 581
Db    628 CGGGCAACTTTATGAGGGGTCTTGAGTCTTCCACAACA-TGGGCAACGGGACT 686
Qy    582 TGACTTGCGTGGCTGGGACGAGGCCGGGACGTGTGGACAGCTGGGACTT 641
Db    687 TGACTTGTGTGGCTGGGACGAGGCCGGAGACTCCTTAACTGTGTCAA-CTGA 745
Qy    642 CTGTGTGATTCGGGTGTCCCC 659
Db    746 CTGTGGTTCAGCTTCCAC 763
```

FEATURES   Location/Qualifiers
     source   1. .794
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4236489"
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP library."

BASE COUNT   186 a   214 c   215 g   179 t
ORIGIN

Query Match   23.7%;   Score 276.4; DB 12;   Length 794;
Best Local Similarity   69.3%;   Pred. No. 1.6e-54;
Matches 405;   Conservative   0;   Mismatches 176;   Indels   3;   Gaps   2;

```
Qy     42 CCGGTTCTGGGTCTGGTAATGAAGTCATAGAAGAGGCCCCAAAATGCAAGAGTCCTGAAGG 101
Db    148 CAGCTTCCGGATCCAGTTATCAGATCATAGAAGTGTAACAGTCCTAAAGG 207
Qy    102 GCTCCCAGGCCTGCTTCAACTGCACCGTCTCCCAGGCCTGGAAGCTCATCATGTGGGCTC 161
Db    208 ACTCAGAGGCTCACTTCAACTGCACCGTGACTCACGGCTGGAAGCTTCTCATGTGGACTC 267
Qy    162 TCAGTGACATGGTGGTGGGTTCAGGCCATGGGAGACCCCATCATCACCAATGACCGGCT 221
Db    268 TTAACCAAAATGGTGGTGCTGAGTCTCACCACCAAGACCATCATCACCAACAACCGCT 327
Qy    222 TCACCTCTCAGAGGTACGACACCAGGGGGGGAACTTCACCTCGGAGATGATCATCACAATG 281
Db    328 TCACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATCCACAATG 387
Qy    282 TGGAGCCCCAGTGATTCGGGGAACATCAGATGCAGCCTTCCAGAACAGCAGTGTCCTGCATGGAT 341
Db    388 TGCAGCCCCAGTGACTCGGGATCCGTGCAATCAGCCCACAGCCGGTTTGGAT 447
Qy    342 CTGCTTACCTTACGGTCGCAAGTGTTATTGGGGAGAGGCGTGTTCATTCCCAGTGTTAATCTTGTAG 401
Db    448 CTGCCTTCCTCTCAGTGCAAGTCATGGGGACCCTGACACTTCCTAGCAACAACCTTATAG 507
Qy    402 TCGCTGAGAATGAAACCTTGTGAAGTTACTTGTTGCTACCCTCACACTGGACCTGGCTCCGG 461
Db    508 TCACTGAGGGTGAACCCTGTAAATGTGACTTGAACCTCACTCCCGG 567
Qy    462 ATATTTCCTGGGGAGCTCGGGTCTCCTGGTCAGCCATTCAAGCTATTATTTTGTTCCGAGC 521
Db    568 ATATTTCCTGGGAGCTTGAGGTTCCCGTTACAAATTCCTTTCTGGAGC 627
Qy    522 CCAGCGACCTTCAAAGTGCAGTGAGCATCTGGCTCGTGACCCCACAGAGCAAATGGGACTT 581
```

OM nucleic - nucleic search, using sw model
                              GenCore version 5.1.4_p5_4578
                              Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 23:25:02 ; Search time 1894.29 Seconds
                              (without alignments)
                              9985.963 Million cell updates/sec

Title: US-09-729-264-3
Perfect score: 1168
Sequence: 1 agtgatcaatggtggcaggag............gtaatacaactgtagtatag 1168
Scoring table: IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0
Searched: 16154066 seqs, 8097743376 residues.

Total number of hits satisfying chosen parameters: 32308132

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database : EST:*
     1: em_esthba:*
     2: em_esthum:*
     3: em_estin:*
     4: em_estmu:*
     5: em_estov:*
     6: em_estpl:*
     7: em_estro:*
     8: em_htc:*
     9: em_estl:*
    10: gb_est2:*
    11: gb_htc:*
    12: gb_est3:*
    13: gb_est4:*
    14: gb_est5:*
    15: em_estfun:*
    16: em_estom:*
    17: gb_gss:*
    18: em_gss_hum:*
    19: em_gss_inv:*
    20: em_gss_pln:*
    21: em_gss_vrt:*
    22: em_stpl:*
    23: em_gss_fun:*
    24: em_gss_mus:*
    25: em_gss_other:*
    26: em_gss_pro:*
    27: em_gss_rod:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	306	26.2	784	12	BG863468	BG863468 602796941
2	276.4	23.7	794	12	BF784177	BF784177 602108039
3	267.4	22.9	725	12	BG740428	BG740428 602633817
4	256.8	22.0	631	10	BB663870	BB663870 BB663870
5	242	20.7	527	10	BE032610	BE032610 132035 MA
6	193.2	16.5	517	10	BE476432	BE476432 159670 BA

No.	Score	%	DB	ID	Description
7	190.8	16.3	17	AG069679	AG069679 Pan trogl
8	190.8	16.3	17	AG107877	AG107877 Pan trogl
9	185.8	15.9	10	AV601192	AV601192 AV601192
10	185.8	15.6	10	AA265274	AA265274 mx91dd09.r
11	181.8	15.6	10	AW990468	AW990468 uf37d05.y
12	181.8	15.6	12	BG173684	BG173684 6C2336823
13	181.8	15.6	11	AK008060	AK008060 Mus muscu
14	180.2	15.4	9	AI425363	AI425363 mx51dd09.y
15	180.2	15.5	17	BG206666	BG206666 RS26117
16	178.6	15.3	12	BB625217	BB625217 BB625217
17	165	14.1	17	AZ379623	AZ379623 1M0134G22
18	153.2	13.1	17	BE376590	BE376590 6C12226419
19	149	12.8	9	AI788300	AI788300 uk56d01.y
20	147.4	12.6	9	AI790785	AI790785 uk28a12.x
21	145.4	12.4	17	AZ411779	AZ411779 1M0185B04
22	145.1	12.1	12	AGI42221	AGI42221 Pan trogl
23	141.6	11.0	12	BF040046	BF040046 BP250004B
24	128.8	10.6	9	AGI42221	AGI42221 Pan trogl
25	123.2	10.5	17	BB846133	BB846133 BB846133
26	110	9.4	13	BI452873	BI452873 BI452873
27	102.8	8.8	13	BI454276	BI454276 60317002I1
28	100	8.8	14	N47851	N47851 yw9h05.r1
29	100	8.6	17	N93995	N93995 za66f09.r1
30	98.4	8.4	10	AQ403719	AQ403719 HS_5049_A
31	78.6	6.7	9	AI647601	AI647601 uk41f12.x
32	78.6	6.7	9	AI648720	AI648720 uk28a12.x
33	78.6	6.7	9	AI788973	AI788973 uk56d01.x
34	77	6.6	10	BB531388	BB531388 BB531388
35	67.8	5.8	10	AW112084	AW112084 MC7851 mo
36	62.8	5.4	10	BB564363	BB564363 BB564363
37	56	4.8	12	BG445021	BG445021 GA__Ea002
38	52.6	4.5	9	AI756062	AI756062 EtEsTea09
39	50.8	4.3	9	AL514929	AL514929 AL514929
40	50	4.3	9	AV431233	AV431233 AV431233
41	49.8	4.3	9	AA695064	AA695064 GM02059.5
42	49.6	4.2	17	AL232043	AL232043 Tetraodon
43	49.6	4.2	17	AL065607	AL065607 Drosophil
44	49.4	4.2	10	BE017134	BE017134 fk76f06.y
45	48.6	4.2	14	BQ599899	BQ599899 MI-P-E6-a

ALIGNMENTS

RESULT 1
LOCUS   BG863468   784 bp   mRNA   linear   EST 29-MAY-2001
DEFINITION 602796941F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4917803 5',
            mRNA sequence.
ACCESSION   BG863468
VERSION   BG863468.1 GI:14214006
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 784)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10830 row: b column: 12
            High quality sequence stop: 768.
FEATURES   Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

```
                    |||||| | |||||||| || | |||||||||||| || | || ||| ||
Db 121168  AGAGGCTCACCTCAACTGCACTGCACTGTGACTCATGGCCTGA 121205
```

Search completed: April 29, 2003, 14:10:30
Job time : 4555.27 secs

TITLE   Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
JOURNAL   Direct Submission
            Submitted (09-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On May 9, 2002 this sequence version replaced gi:20503171.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
COMMENT   ---------------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ---------------------- Project Information
            Center project name: L20921
            Center clone name: 147_E_11
            ---------------------- Summary Statistics
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 169879 bases at least Q40
            Consensus quality: 172188 bases at least Q30
            Consensus quality: 173133 bases at least Q20
            Insert size: 176000; agarose-fp
            Insert size: 173761; sum-of-contigs
            Quality coverage: 8.3 in Q20 bases; agarose-fp
            Quality coverage: 8.4 in Q20 bases; sum-of-contigs
            ----
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *   1   28835: contig of 28835 bp in length
            *   28836   28935: gap of 100 bp
            *   28936   29562: contig of 627 bp in length
            *   29563   29662: gap of 100 bp
            *   29663   30328: contig of 666 bp in length
            *   30329   30428: gap of 100 bp
            *   30429   32005: contig of 1577 bp in length
            *   32006   32105: gap of 100 bp
            *   32106   33128: contig of 1023 bp in length
            *   33129   33228: gap of 100 bp
            *   33229   34512: contig of 1284 bp in length
            *   34513   34612: gap of 100 bp
            *   34613   36079: contig of 1467 bp in length
            *   36080   36179: gap of 100 bp
            *   36180   37483: contig of 1304 bp in length
            *   37484   37583: gap of 100 bp
            *   37584   39166: contig of 1583 bp in length
            *   39167   39266: gap of 100 bp
            *   39267   41065: contig of 1799 bp in length
            *   41066   41165: gap of 100 bp
            *   41166   43203: contig of 2038 bp in length
            *   43204   43303: gap of 100 bp
            *   43304   47489: contig of 4186 bp in length
            *   47490   47589: gap of 100 bp
            *   47590   50604: contig of 3015 bp in length
            *   50605   50704: gap of 100 bp
            *   50705   56370: contig of 5666 bp in length
            *   56371   56470: gap of 100 bp
            *   56471   64079: contig of 7609 bp in length
            *   64080   64179: gap of 100 bp
            *   64180   72619: contig of 8440 bp in length
            *   72620   72719: gap of 100 bp
            *   72720   81566: contig of 8847 bp in length
            *   81567   81666: gap of 100 bp
            *   81667   96874: contig of 15108 bp in length
            *   96675   96874: gap of 100 bp in length

FEATURES   Location/Qualifiers
     source   1..175861
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-147E11"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature   1..28835
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature   28936..29562
                     /note="assembly_fragment"
     misc_feature   29663..30328
                     /note="assembly_fragment"
     misc_feature   30429..32005
                     /note="assembly_fragment"
     misc_feature   32106..33128
                     /note="assembly_fragment"
     misc_feature   33229..34512
                     /note="assembly_fragment"
     misc_feature   34613..36079
                     /note="assembly_fragment"
     misc_feature   36180..37483
                     /note="assembly_fragment"
     misc_feature   37584..39166
                     /note="assembly_fragment"
     misc_feature   39267..41065
                     /note="assembly_fragment"
     misc_feature   41166..43203
                     /note="assembly_fragment"
     misc_feature   43304..47489
                     /note="assembly_fragment"
     misc_feature   47590..50604
                     /note="assembly_fragment"
     misc_feature   50705..56370
                     /note="assembly_fragment"
     misc_feature   56471..64079
                     /note="assembly_fragment"
     misc_feature   64180..72619
                     /note="assembly_fragment"
     misc_feature   72720..81566
                     /note="assembly_fragment"
     misc_feature   81667..96774
                     /note="assembly_fragment"
     misc_feature   96875..119647
                     /note="assembly_fragment"
     misc_feature   119748..147772
                     /note="assembly_fragment"
     misc_feature   147873..174557
                     /note="assembly_fragment"
     misc_feature   174658..175861
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"

BASE COUNT   48903 a   39540 c   39029 g   46286 t   2103 others
ORIGIN

Qy   53 TTCTGGGGTCTGGTAAGTCATAGAAGGCCCCAGAAGTCCAGAAGTCCTGAAGGGCTC 112
               |||||| || || |||| |||||| ||||||| ||||||||||||||
Db   121108 TTCTGGGATCCCGTTATCAGATCATAGACGGCCCTAAGAGAATATAAACAGTGTCCTTACAGGGCTC 121167

Qy   113 CCAGGCGTCGGCGTTCAACTGCACCGTCTCCCAGGGCTGA 150

Query Match   4.9%; Score 58; DB 2; Length 175861;
Best Local Similarity   74.5%; Pred. No. 0.0043;
Matches   73; Conservative   0; Mismatches   25; Indels   0; Gaps   0;

```
Query Match       8.5%;   Score 100;  DB 9;  Length 199665;
Best Local Similarity  100.0%;  Pred. No. 3.3e-13;
Matches 100;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1076  CCACGCGGGCGTGATCCAACGTTCCACCCAGGGCCAGCCAAAGTCATCCRCAGGCTTCTTTTAATCT  1135
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   17555  CCACGCGGGCGTGATCCAACGTTCCACCCAGGGCCAGCCAAAGTCATCCACCAGGCTTCTTTTAATCT  17614

Qy    1136  GGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG  1175
            ||||||||||||||||||||||||||||||||||||||||
Db   17615  GGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG  17654

RESULT 15
AC120346
LOCUS       AC120346               175861 bp   DNA     linear   HTG 09-MAY-2002
DEFINITION  Mus musculus clone RP23-147E11, WORKING DRAFT SEQUENCE, 22 ordered
            pieces.
ACCESSION   AC120346
VERSION     AC120346.3  GI:20514994
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 175861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Unpublished
REFERENCE   2  (bases 1 to 175861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Illev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Nell,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 175861)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Illev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
            Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Nell,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
```

```
repeat_region   33168..33678
                /evidence=not_experimental
                /rpt_family="THE1B"
repeat_region   34563..34620
                /evidence=not_experimental
                /rpt_type=tandem
                /rpt_unit=tttccagaaaagaaggatacgtacttaggca
                /evidence=not_experimental
                note="homology = 93.10%, score = 21, counts = 2"
repeat_region   complement(35785..35824)
                /evidence=not_experimental
                /rpt_family="MLT2CB"
repeat_region   complement(35924..36345)
                /evidence=not_experimental
                /rpt_family="MLT2CB"
repeat_region   complement(36403..36475)
                /evidence=not_experimental
                /rpt_family="MLT2CB"
repeat_region   complement(36717..36828)
                /evidence=not_experimental
                /rpt_family="MLT2CB"
repeat_region   complement(37064..37364)
                /evidence=not_experimental
                /rpt_family="MER67C"
repeat_region   complement(37440..37501)
                /evidence=not_experimental
                /rpt_family="AluY"
repeat_region   37518..38158
                /note="homology = 98.40%, score = 29, counts = 2"
                /rpt_type=tandem
                /rpt_unit=aaaacacttcacttgttcttcaaaactgctc
                /evidence=not_experimental
repeat_region   complement(38477..39376)
                /evidence=not_experimental
                /rpt_family="L2"
repeat_region   complement(39552..39861)
                /evidence=not_experimental
                /rpt_family="L2"
repeat_region   complement(39884..40122)
                /evidence=not_experimental
                /rpt_family="MLT1D"
repeat_region   complement(40243..40542)
                /evidence=not_experimental
                /rpt_family="MLT1D"
repeat_region   40543..41180
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   complement(42066..42114)
                /evidence=not_experimental
                /rpt_family="L2"
repeat_region   42138..42317
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   complement(42332..42710)
                /evidence=not_experimental
                /rpt_family="L2"
repeat_region   complement(42758..46751)
                /evidence=not_experimental
                /rpt_family="L1PB1"
repeat_region   complement(46792..47084)
                /evidence=not_experimental
                /rpt_family="L1PB1"
repeat_region   47189..47992
                /evidence=not_experimental
                /rpt_family="L1MA"
repeat_region   complement(48002..48132)
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   48905..49310
```

```
Db 120358 ACAACCGCTTCACCTATGCCACAGTTACAACAGCACTGACAGTCTTCATCTCGGAGTTGATCA 120299
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    280 TCCACAATGTGGGAGCCCAGTGATTCGGGGACAATCAGATGCAGCCGTTCCAGAACAGTCGCC  339

Db 120298 TCCATGAATGTGGCAGCCCCAGTGACTTCGGGGATCCGTGCAAATGCAGCCTTGCAGAACAGCCAATG 120239
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    340 TCCATGAATCTGCTTTACCTTTACCGTCCAAGTT  371

Db 120238 GGTTTGGATCTGCCTTCCTTCCTTCAGTGTGCAAGGT 120207
           |||||||||||||||||||||||||||||||
```

RESULT 14
AF064857
LOCUS   AF064857   199665 bp   DNA   linear   PRI 23-JUN-1998
DEFINITION Homo sapiens BAC derived from chromosome 21q22.3, complete
            sequence, containing PEP19 (PCP4) gene.
ACCESSION   AF064857
VERSION   AF064857.1 GI:3171149
KEYWORDS   HTG.
SOURCE   Homo sapiens
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 199665)
  AUTHORS   Taudien,S., Nordsiek,G., Korenberg,J., Drescher,B., Weber,J.,
            Schattevoy,R. and Rosenthal,A.
  TITLE   Direct Submission
  JOURNAL   Submitted (11-MAY-1998) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES   Location/Qualifiers
     source   1..199665
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
     repeat_region   complement(1..383)
                     /evidence=not_experimental
                     /rpt_family="L1"
     repeat_region   747..814
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(1345..1500)
                     /evidence=not_experimental
                     /rpt_family="L1PA4"
     repeat_region   1499..1588
                     /evidence=not_experimental
                     /rpt_family="L1PA2"
     repeat_region   complement(1584..2179)
                     /evidence=not_experimental
                     /rpt_family="L1PA2"
     repeat_region   2174..3405
                     /evidence=not_experimental
                     /rpt_family="L1PA2"
     repeat_region   3388..3454
                     /evidence=not_experimental
                     /rpt_family="L1P"
     repeat_region   3968..4086
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(6245..6668)
                     /evidence=not_experimental
                     /rpt_family="L1ME3A"
     repeat_region   6671..6969
                     /evidence=not_experimental
                     /rpt_family="L1ME3A"
     repeat_region   complement(7057..7376)
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   complement(7387..7491)
                     /evidence=not_experimental
                     /rpt_family="MER2"
     repeat_region   7781..8020
                     /evidence=not_experimental
                     /rpt_family="L1ME3A"
     repeat_region   8650..8823
                     /evidence=not_experimental
                     /rpt_family="MER58A"
     repeat_region   complement(10456..10720)
                     /evidence=not_experimental
                     /rpt_family="L1PA5"
     repeat_region   complement(18994..19347)
                     /evidence=not_experimental
                     /rpt_family="L2"
     repeat_region   complement(13514..13814)
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(13823..14116)
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(15662..15785)
                     /evidence=not_experimental
                     /rpt_family="AluJo"
     repeat_region   complement(20371..20556)
                     /rpt_unit=agtatctattcctgattggcagtgctcatac
                     /rpt_type=tandem
                     /evidence=not_experimental
                     /rpt_family="MER5A"
     repeat_region   19378..19470
                     /note="homology = 81.70%, score = 28, counts = 3"
                     /evidence=not_experimental
                     /rpt_family="THE1B"
     repeat_region   21213..21500
                     /evidence=not_experimental
                     /rpt_family="MER3"
     repeat_region   22291..22567
                     /evidence=not_experimental
                     /rpt_family="AluSx"
     repeat_region   complement(22583..23201)
                     /evidence=not_experimental
                     /rpt_family="MER39b"
     repeat_region   23202..23471
                     /evidence=not_experimental
                     /rpt_family="L1PA2"
     repeat_region   24223..24454
                     /evidence=not_experimental
                     /rpt_family="MER39b"
     repeat_region   24455..24889
                     /evidence=not_experimental
                     /rpt_family="MSTB-internal"
     repeat_region   complement(25434..25767)
                     /evidence=not_experimental
                     /rpt_family="MSTB"
     repeat_region   25993..26294
                     /evidence=not_experimental
                     /rpt_family="MER58B"
     repeat_region   27222..27283
                     /evidence=not_experimental
                     /rpt_family="AluSq"
     repeat_region   complement(28932..29335)
                     /evidence=not_experimental
                     /rpt_family="MIR"
     repeat_region   complement(29462..30001)
                     /evidence=not_experimental
                     /rpt_family="MSTA"
     repeat_region   complement(30308..30676)
                     /evidence=not_experimental
                     /rpt_family="MER34"
     repeat_region   31012..31364
                     /evidence=not_experimental
                     /rpt_family="MLT1B"
     repeat_region   complement(32507..32738)
                     /evidence=not_experimental
                     /rpt_family="MLT1A2"
     repeat_region   complement(32766..32980)
                     /evidence=not_experimental
                     /rpt_family="MSTD"
```

```
-------------- Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
             Center project name: L20921
             Center clone name: 147_E_11
-------------- Summary Statistics
             Sequencing vector: Plasmid; n/a; 100% of reads
             Chemistry: Dye-terminator Big Dye; 100% of reads
             Assembly program: Phrap; version 0.960731
             Consensus quality: 169879 bases at least Q40
             Consensus quality: 172188 bases at least Q30
             Consensus quality: 173133 bases at least Q20
             Insert size: 176000; agarose-fp
             Insert size: 173761; sum-of-contigs
             Quality coverage: 8.3 in Q20 bases; agarose-fp
             Quality coverage: 8.4 in Q20 bases; sum-of-contigs
--------------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
*        1    28835: contig of 28835 bp in length
*    28836    28935: gap of    100 bp
*    28936    29562: contig of   627 bp in length
*    29563    29662: gap of    100 bp
*    29663    30328: contig of   666 bp in length
*    30329    30428: gap of    100 bp
*    30429    32005: contig of  1577 bp in length
*    32006    32105: gap of    100 bp
*    32106    33128: contig of  1023 bp in length
*    33129    33228: gap of    100 bp
*    33229    34512: contig of  1284 bp in length
*    34513    34612: gap of    100 bp
*    34613    36079: contig of  1467 bp in length
*    36080    36179: gap of    100 bp
*    36180    37483: contig of  1304 bp in length
*    37484    37583: gap of    100 bp
*    37584    39166: contig of  1583 bp in length
*    39167    39266: gap of    100 bp
*    39267    41065: contig of  1799 bp in length
*    41066    41165: gap of    100 bp
*    41166    43203: contig of  2038 bp in length
*    43204    43303: gap of    100 bp
*    43304    47489: contig of  4186 bp in length
*    47490    47589: gap of    100 bp
*    47590    50604: contig of  3015 bp in length
*    50605    50704: gap of    100 bp
*    50705    56370: contig of  5666 bp in length
*    56371    56470: gap of    100 bp
*    56471    64079: contig of  7609 bp in length
*    64080    64179: gap of    100 bp
*    64180    72619: contig of  8440 bp in length
*    72620    72719: gap of    100 bp
*    72720    81566: contig of  8847 bp in length
*    81567    81666: gap of    100 bp
*    81667    96774: contig of 15108 bp in length
*    96775    96874: gap of    100 bp
*    96875   119647: contig of 22773 bp in length
*   119648   119747: gap of    100 bp
*   119748   147772: contig of 28025 bp in length
*   147773   147872: gap of    100 bp
*   147873   174557: contig of 26685 bp in length
*   174558   174657: gap of    100 bp
*   174658   175861: contig of  1204 bp in length.
```

FEATURES   Location/Qualifiers

```
source          1..175861
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /clone="RP23-147E11"
                /clone_lib="RPCI-23 Female Mouse BAC"
                1..28835
                /note="assembly_fragment
                clone_end:SP6
                vector_side:left"
misc_feature    28936..29562
                /note="assembly_fragment"
misc_feature    29663..30328
                /note="assembly_fragment"
misc_feature    30429..32005
                /note="assembly_fragment"
misc_feature    32106..33128
                /note="assembly_fragment"
misc_feature    33229..34512
                /note="assembly_fragment"
misc_feature    34613..36079
                /note="assembly_fragment"
misc_feature    36180..37483
                /note="assembly_fragment"
misc_feature    37584..39166
                /note="assembly_fragment"
misc_feature    39267..41065
                /note="assembly_fragment"
misc_feature    41166..43203
                /note="assembly_fragment"
misc_feature    43304..47489
                /note="assembly_fragment"
misc_feature    47590..50604
                /note="assembly_fragment"
misc_feature    50705..56370
                /note="assembly_fragment"
misc_feature    56471..64079
                /note="assembly_fragment"
misc_feature    64180..72619
                /note="assembly_fragment"
misc_feature    72720..81566
                /note="assembly_fragment"
misc_feature    81667..96774
                /note="assembly_fragment"
misc_feature    96875..119647
                /note="assembly_fragment"
misc_feature    119748..147772
                /note="assembly_fragment"
misc_feature    147873..174557
                /note="assembly_fragment"
misc_feature    174658..175861
                /note="assembly_fragment
                clone_end:T7
                vector_side:right"
BASE COUNT     48903 a   39540 c   39029 g   46286 t    2103 others
ORIGIN
```

```
Query Match           15.1%;     Score 177.6;  DB 2;   Length 175861;
Best Local Similarity  72.9%;     Pred. No. 6.5e-32;
Matches  242;  Conservative  0;  Mismatches  89;  Indels  1;  Gaps  1;

Qy      41 TTTCTTCCACGGTTCTTGGGTCTGTGTAATGAAGTCATTAGAAGGCCCCCAGAAATGCAACAGT 100
            ||| || |||| | || ||||  ||| ||||||||||||||||||||| || |||||||||| |
Db  120538 TTGCCTTCTTAGCTTCCGGATCCAGTTATCAGATCATTAGAAGGTCCTCAGAATGTAACAGT 120479

Qy     101 CCTGAAGGGGCTCCAGGCTCGGCTTCAACTGCACCGT-CTCCCACGGCGTGGAAGCTCATCA 159
            |||  ||||||||||| |||||||||||||||||||| | |||||||||||||||||||| ||
Db  120478 CCTAAAGGGACTCAGAGGCTCACTCAACTGCACCGGTGGGCTCACGGCTGGAAGCTTCTCA 120419

Qy     160 TGTGGGCTCTCAGTGACATGGTGTGCTAAGCGTCAGCGGCCCATGAGCGCCATCATCACCA 219
            ||| |||||||||||| |||| | ||||| | |||||||||||||||||||||||||||||
Db  120418 TGTGGGACTCTTAACCAAAATGGTGGTGCTCAGTCTCACCACCAAGGAGACCCATCATCACCA 120359

Qy     220 ATGACCGCGTTCACCTCTTCAGAGGTACGACCAGGGCGGGGGGAACTTCACCTCGGAGATGATCA 279
```

COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
               Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H.; Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 10 Row: m Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.

FEATURES   Location/Qualifiers
     source   1..1556
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="FVB/N"
                     /clone="MGC:7960 IMAGE:3584645"
                     /tissue_type="Mammary tumor. Metallothionien-TGF alpha
                     model. 10 month old virgin mouse. Taken by biopsy."
                     /clone_lib="NCI_CGAP_Mam1"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS   55..867
                     /codon_start=1
                     /product="RIKEN cDNA 2010003D20 gene"
                     /protein_id="AAH04806.1"
                     /db_xref="GI:13435933"
                     /db_xref="LocusID:72058"
                     /translation="MEGSWRDVLAVLVILAQLIASGSSYQIIEGPQNVTLKDSEAHF
                     NCTVTHGWKLLMWTLNQMVVLSLTTQGPIINNRFTYASYNSTDSFISELIIHDVQPS
                     DSGSVQSLQNSHGFGSAFLSVQDSIGEEGPALPTWAIILLAVAFSLLLIIVLLII
                     FCCCASREKEESTYQNEIRKSANMRTNKADPETKLKSGKENYGYSSDEAKAAQTAS
                     LPPKSAEVSLPEKRSSSLPYQELNKHQPGPATHPRVSFDIASPQKYRNVTLV"
BASE COUNT   447 a   364 c   367 g   378 t
ORIGIN

        Query Match   15.8%;   Score 186;   DB 10;   Length 1556;
        Best Local Similarity   71.1%;   Pred. No. 4.6e-34;
        Matches 246;   Conservative 0;   Mismatches 100;   Indels 0;   Gaps 0;

Qy   24 GAGATGGGGGCTTGTGATTTTCCTCCACGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGC   83
             | || |||| || |||| | |||| | || |||||||| ||| |||||||||
Db   85 GTGCTGGGTCATCCTGCTCAGCTGACCAGCTTCCGGATCCAGTTATCACGATCAATGAAGGT   144

Qy   84 CCCCAGAATGCAACAGTCCTCTGAGAGGGTCCCCAGCGCTCCCTTCAAACTGCACCGTCTCCCAG   143
             ||||||||| || |||| | ||| || || ||| || || ||| |||||| | ||| || |||
Db   145 CCTCAGAATGTAACAGTCCTAAAGGACTCAGAGGCTCACTTCAACTGCACCGTCGACTCAC   204

Qy   144 GGGCTGGAAGCTCATCATGTGGGGTGTGGGGTTGTGGTGCTAAGCGTCAGGCCCATG   203
             |||||||||| ||||| ||| ||| |||| ||||| | |||| || ||||| |||
Db   205 GGGCTGGGAAGCTTCTCATGTGGGGACTTCTTAACCAAAATGGGTGGGTGCTCACCACCAA   264

Qy   204 GAGGCCCATCATCACCAATGACCGCTTCACCCTCTCAGAGGTACGACCAGGGCGGGGAACTTC   263
             ||||||| |||| | ||| | || ||||||| ||| ||||| ||| |||| ||| |||| ||
Db   265 GGACCCATCACCAACACCGCCTTCACCTATGCCAGCTACAACGACACTGACAGCTTC   324

Qy   264 ACCTCGGAGAATGACATCATCCACCAATGTGGAGCCGTCAGTGATTCGGGGAACATCAGATGCAG   323
             ||||| ||| | ||| | || |||| ||| | |||| || || | | ||| |||
Db   325 ATCTCGGAGTTGATCATTCATGATGTGCAGCCCAGTGACTCGGGATCCCTGCAATGCAAC   384

Qy   324 CTCCAGAACAGTCGCCTGCATGGATCGTCGCTTACCTTACGGTCCAAG   369
             | ||||| || | |||| | || | || |||| ||| || ||||
Db   385 CTGCAGAACAGCCATGGGTTTGGAATCTGCCTTCCTCTCAGTGCAAG   430

COMMENT

RESULT 10
```
LOCUS       AX380408      754 bp    DNA    linear    PAT 18-MAR-2002
DEFINITION  Sequence 13 from Patent WO0200710.
ACCESSION   AX380408
VERSION     AX380408.1  GI:19575338
KEYWORDS    .
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Welcher,A.A., Sarmiento,U.M., Schultz,H.J., and Chute,H.T.
TITLE       B7-like molecules and uses thereof
JOURNAL     Patent: WO 0200710-A 13 03-JAN-2002;
            Amgen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..754
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             53..724
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB28222.1"
                     /db_xref="GI:19575339"
                     /translation="MEGSWRDVLAVIVILAQLTAGSSSYQIIEGPQNVTVLKDSEAHF
                     NCTVTHGWKLLMWTLNQMVVLSLTTQGPIITNNRFTYASVNSTDSFISELIIHDVQPS
                     DSGSVQCSLQNSHGGSSAFLSVQESTYQNEIRKSANMRTNKADPETKLKSGKENYGYS
                     SDEAKAAQTASLPPKSAEVSLPEKRSSSLPYQELNKHQPGPATIHPRVSFDIASPQKVR
                     NVTLV"
BASE COUNT    220 a    191 c    175 g    168 t
ORIGIN
```

```
Query Match     15.8%;  Score 186;  DB 6;  Length 754;
Best Local Similarity   71.1%;  Pred. No. 4.4e-34;
Matches  246;  Conservative  0;  Mismatches 100;  Indels  0;  Gaps  0;

Qy    24 GAGATGGGGCTTGTGATTTTCCCCACGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGC 83
         |||| ||||||||||||||         |||  |||||   ||||||||||||||||||
Db    83 GTGCTGGGTCATCCTGGCTCAGCTTCCGGATCCAGTTATCAGATCATAGAAGGT 142

Qy    84 CCCAGAGAATGCAAACAGTCCTGAAAGGCGGCTCCCAGGGCTCGCGTTCACCGTCTCCCAG 143
         ||    ||| | | |||  || ||  |||| | |  ||||  |  | ||||| |||  |||
Db   143 CCTCAGAAATGTAACAGTTCCTAAAGGACTCAGAGGCGTCAGGGTCAGGCGCCCATG 202

Qy   144 GGCTGGAAGCTCATCAATGTGGGCGTCTCAGTGACATGGTTGGTGCTAAGCGTCAGGCCAG 203
         |||||||||  || | ||||||  ||||||||||||| || | ||||| || |||||||
Db   203 GGCTGGAAGCTTCTCAAATGTGGGCGTCTCAGTGACTCAGAGGTCGTGGTCTCTACCACCAA 262

Qy   204 GAGCCCATCATCACCAAATGACCGCTTCACCTCTCACCGTGACGACCAGGGCGGGAACTTC 263
         |||||||||||||||||  ||| ||||| ||||  | || | ||||  |  | |||||
Db   263 GGACCCATCATCACCAACACCGCTTCACCTATGCCAGTTGCACAGCACTGACAGCTC 322

Qy   264 ACCTCGGAGATGATCATCCACAAATGTGGAGGCGCCAGTGAATGTGACATCAGATGCAGC 323
         ||| |  |||| |  |||||| | | | |||||  |  || | | | |  |||||||||
Db   323 ATCTCGGAGTTGATCATCCATTGAATGTGCAGCCCAGTGACTGACTCGCGGGATCCGTGCAATGCAGC 382

Qy   324 CTCCAGAACAGTCGCCTGCAAGGGGATCTGCTTACCTTACCGTCCAAG 369
         ||||||||    | ||| | | |||| | | |  |  ||||| ||||||
Db   383 CTGCAGAACAGCCATGGGTTTGGATCTGCCTTCCTCTCAGTGCAAG 428
```

RESULT 11
```
LOCUS       AX380406      895 bp    DNA    linear    PAT 18-MAR-2002
DEFINITION  Sequence 11 from Patent WO0200710.
ACCESSION   AX380406
VERSION     AX380406.1  GI:19575336
KEYWORDS    .
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Welcher,A.A., Sarmiento,U.M., Schultz,H.J., and Chute,H.T.
TITLE       B7-like molecules and uses thereof
JOURNAL     Patent: WO 0200710-A 11 03-JAN-2002;
            Amgen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..895
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             53..865
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB28221.1"
                     /db_xref="GI:19575337"
                     /translation="MEGSWRDVLAVIVILAQLTAGSSSYQIIEGPQNVTVLKDSEAHF
                     NCTVTHGWKLLMWTLNQMVVLSLTTQGPIITNNRFTYASVNSTDSFISELIIHDVQPS
                     DSGSVQCSLQNSHGGSSAFLSVQDSIGEGSPALPTWAIILLAVAFSLLILIIVLIII
                     FCCCASRREKEESTYQNEIRKSANMRTNKADPETKLKSGKENYGYSSDEAKAAQTAS
                     LPPKSAEVSLPEKRSSSLPYQELNKHQPGPATIHPRVSFDIASPQKVRNVTLV"
BASE COUNT    250 a    223 c    212 g    210 t
ORIGIN
```

```
Query Match     15.8%;  Score 186;  DB 6;  Length 895;
Best Local Similarity   71.1%;  Pred. No. 4.4e-34;
Matches  246;  Conservative  0;  Mismatches 100;  Indels  0;  Gaps  0;

Qy    24 GAGATGGGGCTTGTGATTTTCCTCCACGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGC 83
         |||| ||||||||||||||         |||  |||||   ||||||||||||||||||
Db    83 GTGCTGGGTCATCCTGGCTCAGCTTCCGGATCCAGTTATCAGATCATAGAAGGT 142

Qy    84 CCCAGAATGCAAACAGTCCTGAAAGGCGGCTCCCAGGGCTCGCGTTCAACTGCACCGTCTCCCAG 143
         ||    ||| | | |||  || ||  |||| | |  ||||  |  | ||||| |||  |||
Db   143 CCTCAGAATGTAACAGTTCCTAAAGGACTCAGAGGCGTCAGGGTCAGGCGCCCATG 202

Qy   144 GGCTGAAGCTCATCANTGTGGGCGTCTCAGTGACATGGTTGGTGCTAAGCGTCAGGCCAG 203
         |||||||||  || | ||||||  ||||||||||||| || | ||||| || |||||||
Db   203 GGCTGGAAGCTTCTCAAATGTGGGCGTCTCAGTGACTCAGAGGTCGTGGTCTCTACCACCAA 262

Qy   204 GAGCCCATCATCACCAAATGACCGCTTCACCTCTCACCGTGACGACCAGGGCGGGAACTTC 263
         |||||||||||||||||  ||| ||||| ||||  | || | ||||  |  | |||||
Db   263 GGACCCATCATCACCAACACCGCTTCACCTATGCCAGTTGCACAGCACTGACAGCTC 322

Qy   264 ACCTCGGAGATGATCATCCACAAATGTGGAGGCGCCAGTGAATGTGACATCAGATGCAGC 323
         ||| |  |||| |  |||||| | | | |||||  |  || | | | |  |||||||||
Db   323 ATCTCGGAGTTGATCATCCATTGAATGTGCAGCCCAGTGACTGACTCGCGGGATCCGTGCAATGCAGC 382

Qy   324 CTCCAGAACAGTCGCCTGCAAGGGGATCTGCTTACCTTACCGTCCAAG 369
         ||||||||    | ||| | | |||| | | |  |  ||||| ||||||
Db   383 CTGCAGAACAGCCATGGGTTTGGATCTGCCTTCCTCTCAGTGCAAG 428
```

RESULT 12
```
LOCUS       BC004806     1556 bp    mRNA   linear    ROD 07-AUG-2002
DEFINITION  Mus musculus, RIKEN cDNA 2010003D20 gene, clone MGC:7960
            IMAGE:3584645, mRNA, complete cds.
ACCESSION   BC004806
VERSION     BC004806.1  GI:13435932
KEYWORDS    MGC.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1556)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (21-MAR-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
```

```
LOCUS       AC020851              196900 bp    DNA     linear   HTG 15-JUL-2000
DEFINITION  Mus musculus clone RP21-467L12, WORKING DRAFT SEQUENCE, 33
            unordered pieces.
ACCESSION   AC020851
VERSION     AC020851.2  GI:9211212
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 196900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 15, 2000 this sequence version replaced gi:6686457.
            -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 1437240
            Center clone name: RPCI-21_467L12
            -----
            Summary Statistics
            Consensus quality: 164618 bases at least Q40
            Consensus quality: 180887 bases at least Q30
            Consensus quality: 183777 bases at least Q20
            Estimated Insert size: 147000; pulse field gel estimation
            Estimated Insert size: 19370u; pulse field gel estimation
            Quality coverage:  5.67 in Q20 bases; sum-of-contigs estimation
            Quality coverage:  4.3 in Q20 bases; sum-of-contigs estimation.
            NOTE: This is a 'working draft' sequence. It currently
            consists of 33 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            *           1    1541: contig of 1541 bp in length
            *        1542    1641: gap of unknown length
            *        1642    3492: contig of 1851 bp in length
            *        3493    3592: gap of unknown length
            *        3593    4738: contig of 1146 bp in length
            *        4739    4838: gap of unknown length
            *        4839    7248: contig of 2410 bp in length
            *        7249    7348: gap of unknown length
            *        7349    9178: contig of 1830 bp in length
            *        9179    9278: gap of unknown length
            *        9279   11493: contig of 2215 bp in length
            *       11494   11593: gap of unknown length
            *       11594   13607: contig of 2014 bp in length
            *       13608   13707: gap of unknown length
            *       13708   15946: contig of 2239 bp in length
            *       15947   16046: gap of unknown length
            *       16047   18392: contig of 2346 bp in length
            *       18393   18492: gap of unknown length
            *       18493   21231: contig of 2739 bp in length
            *       21232   21331: gap of unknown length
            *       21332   24662: contig of 3331 bp in length
            *       24663   24762: gap of unknown length
            *       24763   27140: contig of 2378 bp in length
            *       27141   27240: gap of unknown length
            *       27241   29950: contig of 2710 bp in length
            *       29951   30050: gap of unknown length
            *       30051   32245: contig of 2195 bp in length
            *       32246   32345: gap of unknown length
            *       32346   35688: contig of 3343 bp in length
            *       35689   35788: gap of unknown length
            *       35789   40387: contig of 4599 bp in length
            *       40388   40487: gap of unknown length
            *       40488   43580: contig of 3093 bp in length
            *       43581   43680: gap of unknown length
            *       43681   46437: contig of 2757 bp in length
            *       46438   46537: gap of unknown length
            *       46538   50184: contig of 3647 bp in length
            *       50185   50284: gap of unknown length
            *       50285   54572: contig of 4288 bp in length
            *       54573   54672: gap of unknown length
            *       54673   60227: contig of 5555 bp in length
            *       60228   60327: gap of unknown length
            *       60328   68024: contig of 7697 bp in length
            *       68025   68124: gap of unknown length
            *       68125   78610: contig of 10486 bp in length
            *       78611   78710: gap of unknown length
            *       78711   88151: contig of 9441 bp in length
            *       88152   88251: gap of unknown length
            *       88252   96257: contig of 8006 bp in length
            *       96258   96357: gap of unknown length
            *       96358  106227: contig of 9870 bp in length
            *      106228  106327: gap of unknown length
            *      106328  115795: contig of 9468 bp in length
            *      115796  115895: gap of unknown length
            *      115896  126363: contig of 10468 bp in length
            *      126364  126463: gap of unknown length
            *      126464  137189: contig of 10726 bp in length
            *      137190  137289: gap of unknown length
            *      137290  146378: contig of 9089 bp in length
            *      146379  146478: gap of unknown length
            *      146479  161686: contig of 15208 bp in length
            *      161687  161786: gap of unknown length
            *      161787  176149: contig of 14363 bp in length
            *      176150  176249: gap of unknown length
            *      176250  196900: contig of 20651 bp in length.
FEATURES             Location/Qualifiers
     source          1..196900
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP21-467L12"
                     /clone_lib="RPCI mouse PAC library 21"
BASE COUNT    51456 a   43823 c   44608 g   53666 t   3347 others
ORIGIN
```

```
Query Match         16.1%;   Score 188.6;   DB 2;   Length 196900;
Best Local Similarity   73.1%;   Pred. No. 1.5e-34;
Matches  242;   Conservative   0;   Mismatches  89;   Indels   0;   Gaps   0;

Qy         41 TTTCCTCCACGGTTCTTGGGTCTTGGGTAATGAAGTCATAGAAGGCCCCAGAATGCAACAGT 100
              |||||||| ||||||||||||||||||||| ||||||||||||||||||||| ||||||||
Db      67436 TTGCCTTCTTAGCTTCCGGATCCAGTTATCAGATCATACAGATAGAAGTTCCTCAGAATGT 67495

Qy        101 CCTGAAAGGGCTCCCAGGCTCGCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCAT 160
              ||||  ||||||||| ||||||||||||||||||||| |||||||||||||||||||| |||
Db      67496 CCTAAAGGACTCAGAGGCTTCAACTGCACCGTGACTTCACGGCTGGAAGCTTCTCAT 67555

Qy        161 GTGGGCTCTCAGTGACATAGGTGGTGGTGCTAAGCGTTCAGGCCCCATGGGCCCATCACCAA 220
              ||||||||||||||| ||||||||||| ||||||||||||||||||||||||||||||||
Db      67556 GTGGACTCTTAACCAAATGGTGGTGTTGGCTCACCGTCTCACCAAGGACCCATCATCACCAA 67615

Qy        221 TGACCGCTTCACCTCTCAGAGGTACGACCCAGGGGCGGGGGAACTTCACCTCGGAGATGATCAT 280
              |||||||||| ||||||||| |||||||||||||||||||| ||||||||||||||||||
Db      67616 CAACCGCTTCACCTTATGCCAGTTACACAGCACACTGACAGCTTCATCTCGGGAGTTGATCAT 67675

Qy        281 CCACAAATGTCGCAGCCCAGTGATTCGGGGGAACATCAGATGCAGCCTCCAGAACAGTCGCCT 340
              || || |||||||||||||||||||||||||| ||||||||||||| |||||||| ||||||
Db      67676 CCATGATGTGCAGCGCCCAGTGACTCGGGATCCGTGCAAATGCAGCCTGCAGAAACAGCCATGG 67735

Qy        341 GCATGGATCTGCTTACCTTACCGTCCAAGTT 371
              ||| || |||||||||||||||||||||||
Db      67736 GTTTGGATCTGCTTCCTCTCAGTGCCAAGGT 67766
```

```
repeat_region
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="21"
               /map="21q22.3"
               /clone="BAC-291B3, 3' partial"
               /clone_lib="BAC library"
               /note="Accession No. AF064857"
repeat_region  1088..1566
               /note="L1M4"
               /rpt_family="LINE/L1"
               /rpt_type=DISPERSED
repeat_region  2398..2495
               /note="(TCTA)n"
               /rpt_family="Simple_repeat"
               /rpt_type=TANDEM
repeat_region  complement(2496..2878)
               /note="THE1C"
               /rpt_family="LTR/MaLR"
               /rpt_type=DISPERSED
repeat_region  2879..2931
               /note="(TCTA)n"
               /rpt_family="Simple_repeat"
               /rpt_type=TANDEM
repeat_region  2932..3004
               /note="(TCCA)n"
               /rpt_family="Simple_repeat"
               /rpt_type=TANDEM
repeat_region  3005..3074
               /note="(TCTA)n"
               /rpt_family="Simple_repeat"
               /rpt_type=TANDEM
gene           3473..19157
               /gene="SH3BGR"
mRNA           join(3473..3565,12649..12678,15331..15460,18739..19157)
               /gene="SH3BGR"
CDS            join(3473..3565,12649..12678,15331..15426)
               /partial
               /gene="SH3BGR"
               /note="Accession No. X93498"
               /codon_start=1
               /product="21-Glutamic Acid Rich protein 21-GARP"
               /protein_id="CAB90445.1"
               /db_xref="GI:7717370"
               /translation="GSEKAEEGGETEAQKEGSEDVGNLPEAQEKNEEEGTATEETEE
               IAMEGAEGEAEEEETEAEGEEPGEDEDS"
exon           3473..3565
               /gene="SH3BGR"
               /number=4
repeat_region  5542..5829
               /note="AluSx"
               /rpt_family="SINE/Alu"
               /rpt_type=DISPERSED
repeat_region  complement(6887..7067)
               /note="L1MD3"
               /rpt_family="LINE/L1"
               /rpt_type=DISPERSED
repeat_region  complement(8482..8762)
               /note="AluJo"
               /rpt_family="SINE/Alu"
               /rpt_type=DISPERSED
repeat_region  complement(8865..8964)
               /note="L2"
               /rpt_family="LINE/L2"
               /rpt_type=DISPERSED
repeat_region  9131..9424
               /note="AluSq"
               /rpt_family="SINE/Alu"
               /rpt_type=DISPERSED
repeat_region  complement(9723..9882)
               /note="L2"
               /rpt_family="LINE/L2"
               /rpt_type=DISPERSED
repeat_region  complement(9989..10077)
               /note="L2"
               /rpt_family="LINE/L2"
               /rpt_type=DISPERSED

repeat_region  complement(10078..10429)
               /note="THE1B"
               /rpt_family="LTR/MaLR"
               /rpt_type=DISPERSED
repeat_region  complement(10430..10651)
               /note="L2"
               /rpt_family="LINE/L2"
               /rpt_type=DISPERSED
repeat_region  10980..11000
               /note="(TTTA)n"
               /rpt_family="Simple_repeat"
               /rpt_type=TANDEM
repeat_region  complement(11002..11282)
               /note="AluSp"
               /rpt_family="SINE/Alu"
               /rpt_type=DISPERSED
repeat_region  complement(11315..11556)
               /note="AluJo"
               /rpt_family="SINE/Alu"
               /rpt_type=DISPERSED
repeat_region  11702..11753
               /note="MIR"
               /rpt_family="SINE/MIR"
               /rpt_type=DISPERSED
exon           12649..12678
               /gene="SH3BGR"
               /number=5
repeat_region  complement(13115..13425)
               /note="AluSx"
               /rpt_family="SINE/Alu"
               /rpt_type=DISPERSED
repeat_region  complement(13598..13892)
               /note="AluY"
               /rpt_family="SINE/Alu"
               /rpt_type=DISPERSED
repeat_region  14013..14431
               /note="(CA)n"
               /rpt_type=DISPERSED
               /rpt_family="Simple_repeat"

Query Match 27.3%;  Score 320.2;  DB 9;  Length 340000;
  Best Local Similarity 97.6%;
  Matches 325;  Conservative 0;  Mismatches 8;  Indels 0;  Gaps 0;
                                  Pred. No. 2.9e-66;

Qy      42 TTCCTCCACGGTTCTGGGTCTCTGGTAATGAAGTCATAGAAGAGCCCCCAGAATGCAACAGTC 101
           ||| |||||||||||||||||| ||||||||| ||||||||||||||||||  |||||| |||
Db  268445 TACCTTCCACGGTTCTGGGTCTTGGTAATGAACTCATAGAAGAGCCCCCAAAATGCAAGAGTC 268504

Qy     102 CTGAAGGGCTCCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAGCTCATCATG 161
           ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db  268505 CTGAAGGGCTCCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATG 268564

Qy     162 TGGGCTCTCAGTGACATGGTGCGTGTGCTTAAGCGTCAGGCCCATGGAGCCCATCATCACCAAT 221
           ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db  268565 TGGGCTCTCAGTGACATGGTGGTGCGTGTGCTTAAGCGTCAGGCCCATGGAGCCCATCATCACCAAT 268624

Qy     222 GACCGGCTTCACCTCTCGAGAGGTACCACCAGGCGCGGGAACTTCACCTCCGAGATGATCATC 281
           ||||  |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  268625 GACCCGCTTCACCTCTCGAGAGGTACCACCAGGCGCGGGAACTTCACCTCGGAGATGATCATC 268684

Qy     282 CACAATGTGGAGCCCAGTGATTCGGGGAACATCAGATGCAGCCTCCAGAACAGTCGCCTG 341
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  268685 CACAATGTGGAGCCCAGTGATTCGGGGAACATCAGATGCAGCCTCCAGAACAGTCGCCTG 268744

Qy     342 CATGGATCTGCTTACCTTACCGTCCAAGGTATG 374
           ||||||||||||||||||||||||||||||| ||
Db  268745 CATGGATCTGCTTACCTTACCGTCCAAGGTGTG 268777

RESULT 9
AC020851
```

```
     repeat_region   /rpt_family="MER34"
                     complement(17633..17969)
                     /evidence=not_experimental
     repeat_region   /rpt_family="AluSg1"
                     complement(18004..18334)
                     /evidence=not_experimental
     repeat_region   /rpt_family="LTR17"
                     complement(18335..18632)
                     /evidence=not_experimental
     repeat_region   complement(18335..18632)
                     /rpt_family="HERV17"
                     18352..18420
                     /note="GRAIL, score = 51.000%, comment = good shadow"
                     /evidence=not_experimental
     exon            complement(18397..18584)
                     /note="GRAIL, score = 66.000%, comment = good"
                     /evidence=not_experimental
     exon            complement(18643..24647)
                     /evidence=not_experimental
     repeat_region   complement(19228..19361)
                     /rpt_family="HERV17"
                     /note="MZEF, score = 51.4%"
                     /evidence=not_experimental
     exon            21150..21275
                     /note="MZEF, score = 92.8%"
                     /evidence=not_experimental
     exon            20900..21135
                     /note="MZEF, score = 51.4%"
                     /evidence=not_experimental
     exon            20900..21135
                     /note="Xpound exon prediction, score = 74% (0%)"
                     /evidence=not_experimental

Query Match     27.3%;  Score 320.2;  DB 9;  Length 142742;
Best Local Similarity   97.6%;  Pred. No. 2.8e-66;
Matches 325;  Conservative  0;  Mismatches  8;  Indels  0;  Gaps  0;

Qy      42 TTCCTCCACGGTTGTGGGTCTTGGTAATGAAGTCATAGAAGGCCCCCCAGAATGCAACAGTC 101
           | ||| || |||||| ||||||||||||||||||||||||||||||||||||  |||||||||
Db   44265 TACCTTCCAGGTTCTGGGTTCTTGGTAATGAAGTCATAGAAGGCCCCCCAGAAAATGCAACAGTC 44324

Qy     102 CTGAAGGGCTCCCAGGCTCGCGCTTCAACTGCACCGTCTCCCAGGCGTGGAAGCTCATCATG 161
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   44325 CTGAAGGGCTCCCAGGCTCGCGCTTCAACTGCACCGTCTCCCAGGCGTGGAAGCTCATCATG 44384

Qy     162 TGGGCTCTCAGTGACACATGGTGTGTTGTCTAAGGCCTTCAGCCCCCATGGACCCCATCATCCACCAAT 221
           |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   44385 TGGGCTCTCAGTGACACATGGTGTGTCTAAGGCCTTCAGCCCCCATGGACCCCATCATCACCAAT 44444

Qy     222 GACCGCTTCACCTCTCAGAGGTACGACCAGGCGGGGAACTTCACCTCGGAGATGATGATCATC 281
           |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   44445 GACCGCTTCACCTCTCAGAGTACGACCAGGCGGGGAACTTCACCTCGGAGATGATGATCATC 44504

Qy     282 CACAATGTGGAGCGCCAGTGAATCGGGGAACATCAGATGCAGCCTCAGAACAGTCGCCTG 341
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   44505 CACAATGTGGAGCGCCAGTGAATCGGGGAACATCAGATGCAGCCTCAGAACAGTCGCCTG 44564

Qy     342 CAIGGGATCTGCTTACCTTACCGTCCAAGTTATG 374
           || ||||||||||||||||||||||||||| | |
Db   44565 CATGGGATCTGCTTACCTTACCGTCCAAGGTGTG 44597
```

RESULT 8

```
repeat_region   complement(140..415)
                /evidence=not_experimental
                /rpt_family="MLT2E"
repeat_region   complement(668..881)
                /evidence=not_experimental
                /rpt_family="MER58A"
exon            complement(1078..1162)
                /evidence=not_experimental
                /note="GRAIL, score = 86.000%, comment = excellent shadow"
exon            1114..1403
                /evidence=not_experimental
                /note="GRAIL, score = 86.000%, comment = excellent shadow"
repeat_region   1481..1603
                /evidence=not_experimental
                /rpt_family="MLT1J"
repeat_region   1576..1710
                /evidence=not_experimental
                /rpt_family="L2"
exon            1576..1710
                /evidence=not_experimental
                /note="MZEF, score = 66.18"
exon            complement(1693..1741)
                /evidence=not_experimental
                /note="GRAIL, score = 99.000%, comment = excellent"
repeat_region   complement(1780..2326)
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   3012..3122
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   complement(1834..1955)
                /evidence=not_experimental
                /rpt_family="MLT1E"
exon            2629..2831
                /evidence=not_experimental
                /note="GRAIL, score = 50.000%, comment = good shadow"
repeat_region   3123..3478
                /evidence=not_experimental
                /rpt_family="MLT1J"
repeat_region   3479..3618
                /evidence=not_experimental
                /rpt_family="THE1B"
exon            complement(3502..3540)
                /evidence=not_experimental
                /note="GRAIL, score = 65.000%, comment = good shadow"
exon            4674..4781
                /evidence=not_experimental
                /note="GRAIL, score = 42.000%, comment = marginal"
exon            4712..4749
                /evidence=not_experimental
                /note="Xpound exon prediction, score = 70% (0%)"
repeat_region   5385..6104
                /evidence=not_experimental
                /rpt_family="HERVL"
exon            6682..6751
                /evidence=not_experimental
                /note="MZEF, score = 95.18"
exon            6746..6751
                frame: 1 phase: 0"
                /note="GenScan, score = 1.83%, comment = Initial_exon 6 bp
repeat_region   6901..7339
                /evidence=not_experimental
                /rpt_family="MLT1C"
repeat_region   7347..7457
                /evidence=not_experimental
                /rpt_family="LTR16C"
repeat_region   complement(7561..7696)
                /evidence=not_experimental
                /rpt_family="MER34"
repeat_region   complement(7819..8220)
                /evidence=not_experimental
                /rpt_family="THE1A"
```

```
exon            8097..8119
                /evidence=not_experimental
                /note="GRAIL, score = 45.000%, comment = marginal"
repeat_region   complement(8738..9232)
                /evidence=not_experimental
                /rpt_family="MLT1C"
exon            complement(9273..9420)
                /evidence=not_experimental
                /note="GRAIL, score = 71.000%, comment = good"
exon            9290..9371
                /evidence=not_experimental
                /note="GRAIL, score = 81.000%, comment = excellent shadow"
repeat_region   9966..10041
                /evidence=not_experimental
                /note="homology = 100.000%, score = 38, counts = 2"
                /rpt_type=tandem
                /rpt_unit=ttcttgcacacttcctctgttgagaggtcggtttatcgc
repeat_region   10244..10664
                /evidence=not_experimental
                /rpt_family="MLT1F"
repeat_region   10665..11028
                /evidence=not_experimental
                /rpt_family="MLT1F"
repeat_region   11318..11417
                /evidence=not_experimental
                /rpt_family="T1MC/D"
repeat_region   11503..11818
                /evidence=not_experimental
                /rpt_family="MLT1C"
repeat_region   12502..13180
                /evidence=not_experimental
                /rpt_family="AluSg"
exon            complement(13057..13161)
                /evidence=not_experimental
                /note="MZEF, score = 76.18"
repeat_region   complement(13226..13364)
                /evidence=not_experimental
                /rpt_family="L1MC2"
repeat_region   complement(13613..13908)
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   complement(14317..14614)
                /evidence=not_experimental
                /rpt_family="LTR37A"
exon            complement(14962..15036)
                /evidence=not_experimental
                /note="MZEF, score = 93%"
repeat_region   15255..15299
                /evidence=not_experimental
                /rpt_family="L1PA6"
exon            complement(15469..15497)
                /evidence=not_experimental
                /note="Xpound exon prediction, score = 628 (0%)"
repeat_region   16045..16768
                /evidence=not_experimental
                /rpt_family="MER21A"
exon            complement(16555..16555)
                /evidence=not_experimental
                /note="GRAIL, score = 54.000%, comment = good"
repeat_region   16823..16850
                /evidence=not_experimental
                /rpt_family="MER34"
repeat_region   complement(16851..17208)
                /evidence=not_experimental
                /rpt_family="THE1A"
repeat_region   17209..17632
                /evidence=not_experimental
```

```
LOCUS       AX380404                1195 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 9 from Patent WO0200710.
ACCESSION   AX380404
VERSION     AX380404.1  GI:19575334
KEYWORDS    .
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
TITLE       B7-like molecules and uses thereof
JOURNAL     Patent: WO 0200710-A 9 03-JAN-2002;
            Amgen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1195
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             53..1165
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="CAD28220.1"
                     /db_xref="GI:19575335"
                     /translation="MEGSWRDVLAVIVLIVIILAQLTASSSSYQIIEGPQNVTLKDSEAHF
                     NCTVTHGNKLLMWTLNQMVVLSLTTQPIIINNRFTVASYNSTDSFISELIIHDVQPS
                     DSGSVQCSLQNSHGFGSAFLSVQMGTLNIPSNNLIVTEGEPCNVTCYAVGWTSLPDI
                     SWELEVSSVSHSSYNSTEPGNEMRVLSVLDLIPLGHGTTCLVAELKDLQASKLSIVNL
                     TVQQPPPDSIGEEGPALPTWAIILLAYAFSLLLLILILIVLIIFCCCASRREKEESTY
                     QNEIRKSANMRINKADPETKLKSGKENYGSSDEARAAQTASLPPRSAEVSLPEKRSS
                     SLPYQELNKHQPGPATHPRVSFDIASPQKVRNVTLV"
BASE COUNT      313 a    304 c    291 g    287 t
ORIGIN
```

```
Query Match             32.4%;  Score 381;  DB 6;  Length 1195;
Best Local Similarity   61.9%;  Pred. No. 4.6e-81;
Matches   712;  Conservative    0;  Mismatches  370;  Indels   69;  Gaps    4;

Qy    24 GAGATGGGGCGGCTTGGTGATTTTCCTCCACGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGC 83
           ||| |||| ||    ||||  ||||| | ||||| | |||||||| | |||||||||||||||
Db    83 GTGCTGGTCTCATCCTGGCCTCAGCGTGACAACTTCACCCGGATCCAGTTATCAGATCATAGAAGGT 142

Qy    84 CCCCAGAATGCAACAGTCCTGAAGGGTCCCAGCGTTCCCAACCCTGCCTTCAACACCCTCCCAG 143
           |||||||| | || ||  || ||||  |||||   |||||  |||  |||  || ||| |||||
Db   143 CCTCAGAATGTAACAGTGCTAAAGGACTCACTTCAAACTCGCACCGTGACTCAC 202

Qy   144 GGGCTGGGAAGCTCATCATGTGGGCGCTCTCAGTGACATGGTGGGTCCTAAGCGTCAGGCCCATG 203
           |||||||| |||| |   ||||||| |||||||||| ||| ||||| || ||||||||||||||
Db   203 GGGCTGGGAAGCTTCTCATGGGAGACTCTTAACCAAAATGGTGGGTGCTCACCACCACCAA 262

Qy   204 GAGCCCCATCATCACCAATGACCGGCTTCACCCTCTCAGAGGTACAGACCAGGGCGGGAACTTC 263
           |||||| ||||||||||| ||  | |||||||| |||||| |||   |||||| | ||||| | |
Db   263 GGACCCATCATCACCAACAACCGCTTCACCCTATGCCAGTTACAACAGCACTCAACAGCTTC 322

Qy   264 ACCTCGGAGATGAGATCACCAAATGTGGAGCCCCAGCTGATTCGGGGACAATCAGATGCAGC 323
           ||| || | || | || || |||||||||||||| | |||  || | ||  |||||| ||||
Db   323 AATCTCGGAGTTGATCATCACCATGAGTGCACCACCATCAGGACTCGGGATCCGTGCAATGCAGC 382

Qy   384 TTCATTCCCAGTGTTAATCTTGTAGTCGCTGAGAAATGAACCTTGTGAAGTTACTTGTCTA 443
           || |||||| || | ||   || |  |  | || || ||   ||||| | | ||  |||||
Db   383 CTGCAGAACAGCCATGGTTGGATCGTTGCCCTTCCTCAGTGCAAGTCATCATGGGGACCCTG 442

Qy   444 CCCTCACACTGGACCCGGCTCCCGGATATTCCTCGGTGTCTTCCTGGTCACGCCAT 503
           |||||||| ||| |||  | | |||| ||  ||  ||||  |||| | |||||||
Db   443 AACATTCCTAGCACAACAACCTTATAGTCACTGAGGGTGCAACCCTGTTGACTTCGACTCCCAT 502

Qy   504 TCAAGCTATTATTTGTTCCGGAGCCCAGCGCAACTTCAAAGTGCACTGAGCATCCTGGCT 563
           |||| || |  | ||  |||||||  ||| | |||| | |  | ||| |||  |||||||
Db   503 GCCGTGGGCCTGTGACCTCACTCCCCGGATATTTCCTGGGAGCTTGAGGTTCCGGTAGCCAT 562

Qy   563 TCGAGTTACAAATTCCTTTCTGGAGCCGCGGGCAACTTTATGAGGGTCTTGAGTCTCTGGAC 622
           |||||||| |  |  |   | | ||||  | |||||||| |||| |  |||||||||||||||
Db   563 TCGAGTTACAAATTCCTTTCTGGAGCCGCGGGCAACTTTATGAGGGTCTTGAGTCTCTGGAC 622
```

```
RESULT 7
AF121782

Qy   564 CTGACCCCACGAGCCAATGGGAGTTTTGACTTTGGTGGTCTGGTAGGCGTTCCAGCC 623
           || ||| || |||| || || ||||| |||||| ||  ||||  |||| ||||||||
Db   623 CTCACCACCACTGGGCCAACCGGACCCTTGACTTGTGTGGCGAGAGCTGAAGGACTTGCCAGCC 682

Db   683 AGCAAGTCCTTAACTGTCGAACTGTTGGTTCAGCCCTTCACCTGACCAGTGATTGGAGAG 742
           |||| |||| | || |  | ||||||| | ||||| || || |||| | |||| | |||||
Qy   684 ATTGAATATTTCCAGGTGTTATTATCCAAGTTTACCGAGTTTAGTTTTTTTCATTGGCTACTTGG 743

Db   743 GA----------------------------------AGGCCCAGCAACTGCCGCACCTGG 766
           ||                                   |||  | || | ||| ||||||
Qy   744 GGCAAAGTTGGACTTGGACTAGCAGGCACCAGCTTCTGACGCCGCGACGTGACTCTTACA 803

Qy   804 ATACGCTGCTGCTTGCCGCGGTCGTTGTGGGCTGGCAACTGCTGCTCCGGTTGGTTGT 863
           || ||||| | ||| || ||||  || ||| | || ||||||| || |   || || |
Db   827 ATAATATTCTGTTGCTGTGTT------------GCCTCCAGGAGGAAAAGGAAGAAATC 873

Qy   864 TTCTGCTGTAGAGAAGAAAGAGGATTTCGTATTCATTCAAATTCAAAAGAAATCTGAAAAGAG 923
           ||||| | | ||  || || | |  | || ||| || || ||  |||| | |||||| | |||
Db   874 TACTTATCTAAAATGAAATAAGGAAATTCTGCAAACATGAGGACAA-----------AC 919

Qy   924 AAGACAAAACAAAGAAAATGAGACAGAAAGTGAAAATGAAAACTCCGGCTACAAATTCAGAT 983
           || ||||  ||| || || || |||| | |||  | | ||||||||  || || || |||||
Db   920 AAAGCAGATCCGGAGCAGACAAAGTTAAAAGGAAGAAAACTACGGGTACAGTTCAGAAATGTG 979

Qy   984 GAACAAAAGACACACAGACACCGCTTCTCTCCCCAAATCCTGTGAATCACTGATCCT 1043
           || | | || ||| ||| || | ||| ||||| ||||| | ||| | |||  ||||
Db   980 GAGGCAAAGGCTGCACACAGACACTGCATCTCCCTCCTAAATGTGCTGAAGTCAGCCTTCA 1039

Qy  1044 GAACAAAGAAAACAGTAGCTGTGGGCCCCTCCTCACCACGCGGGGCTGATCAACGGTCCACCCAGG 1103
           |||| | |||  |||| | | |  | | ||  || ||||| ||| |   ||| ||  ||| |||||
Db  1040 GAAAAACCGCAGCAGTA------GCCTTCCTTATCAGGAACTCAATAAACATCAGCCCGGT 1093

Qy  1104 CCAGCAAGTCATCCACACAGGCCTTCTTTTAATCTGGCCAGTCCTGAGAAGGTCAGTAGT 1163
           |||| ||| ||||||||  ||| ||||| ||| | || | |||| || | || |  |  |
Db  1094 CCAGCAACTCATCCACACGGGTTTCTTTGACATCGCCAGTCCTCAGAAGGTCAGAAATGTG 1153

Qy  1164 ACTGTAGTATA 1174
           || |||||| |
Db  1154 ACTTTAGTGTA 1164
```

```
LOCUS       AF121782              142742 bp    DNA     linear   PRI 02-FEB-1999
DEFINITION  Homo sapiens chromosome 21q22.3 PAC 206A10, complete sequence.
ACCESSION   AF121782
VERSION     AF121782.1  GI:4210991
KEYWORDS    HTG.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 142742)
AUTHORS     Taudien,S., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
            Schattevoy,R., Weber,J., Schilling,M., Menzel,U., Yaspo,M.L. and
            Rosenthal,A.
TITLE       Direct Submission
JOURNAL     Submitted (21-JAN-1999) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1..142742
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="PAC 206A10"
     repeat_region   3..118
                     /evidence=not_experimental
                     /rpt_family="MIR"
```

```
RESULT 5
AX380402
LOCUS       AX380402                1139 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 7 from Patent WO0200710.
ACCESSION   AX380402
VERSION     AX380402.1  GI:19575332
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE     B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 7 03-JAN-2002;
            Amgen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1139
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             1..1134
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="CAD28219.1"
                     /db_xref="GI:19575333"
                     /translation="MVAGMENRDPPGPGSSGNEVIEGPQNARVLKGSQARFNCTVSQG
                     WKLIMWALSDMVVLSVRPMEPIITNDRFTSQRDQGNLISEMIIHNVEPSDSGNIRC
                     SLQNSRLHGSAYITVQVWGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWEGLL
                     VSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTCVATWKSLRARKSATVNLTVIRCPQ
                     DTGGGINIPGVLSLSLPSLGFSILPTWGKVGLGLAGTMLLTPTCTITIRCCCCRRCCG
                     NCCCRCCRRKRGNLKKRRQTKKLRQKVEMKTPAITQMNKRPQTPLLISLNPNVPNV
                     ILNKETVAVLLTSGLIINVHPGQQVIHRLLLINPVLRSVIQL"
BASE COUNT      290 a    300 c    283 g    266 t
ORIGIN
```

```
Qy          104 GAAGGCTCCCAGGCTGCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTG 163
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           30 CCCACCCGGTTCTGGGTCTCGGTTAATGAAGTCATAGAAGCGCCCCCAAAAATGCAAGAGTCCT 89

Qy           44 CCTCCACGGTTCTGGGTCTCGGTAATGAAGTCATAGAAGGCCCCCCAGAAATGCAACAGTCT 103
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           30 CCCACCCGGTTCTGGGTCTCGGTTAATGAAGTCATAGAAGCGCCCCCAAAAATGCAAGAGTCCT 89

Query Match             90.8%;  Score 1067.4;  DB 6;  Length 1139;
Best Local Similarity   97.5%;  Pred. No. 1.7e-246;
Matches 1103;  Conservative   0;  Mismatches   6;  Indels  22;  Gaps   1;
```

```
Qy          757 AGCAGGACCACCATCGTTCTTGACGGCGGAGCTGTACTCTTACAAATACGCTGTGCTCTGCCG 816
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           90 GAAGGGCTCCCAGGCTGCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTG 149

Qy          824 CCGTGCGTTGTTGTGGGCGTGCACCAGCTGCTGCGTGTTGTTTCTGCGTGTAGAGAGAAAAG 883
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          164 GGCTCTCCAGTGACATGGTGGTGCTAAGCGTCAGGCCCATGGAGCCGATCATCACCAATGA 223

Qy          824 CCGTGCGTTGTTGTGGGCGTGCACCAGCTGCTGCGTGTTGTTTCTGCGTGTAGAGAGAAAAG 883
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          150 GGCTCTCAGTGACATGGTGGTGCTAAGCGTCAGGCCCATGGAGCCGATCATCACCAATGA 209

Qy          817 CCGTGCGTTGTTGTGGGCTGCCAACTGCGTCCTG 876
                |||||||||||||||||||||||||||||||
Db          210 CCGCTTCACCTCTCAGAGGTACGACCAGGGCGGAAACCTCACCTCGGAGATGATCATCCA 269
```

RESULT 4

```
LOCUS       AX380398        1168 bp    DNA        linear   PAT 18-MAR-2002
DEFINITION  Sequence 3 from Patent WO0200710.
ACCESSION   AX380398
VERSION     AX380398.1  GI:19575328
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE     B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 3 03-JAN-2002;

FEATURES             Location/Qualifiers
     source          1..1168
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             8..1168
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="CAD28217.1"
                     /db_xref="GI:19575329"
                     /translation="MVAGAMENRDPPGSGSGNEVIEGPONARVLKGSQARFNCTVSQG
                     VSHSSYFVPEPSDLQSAVSILALIPQSNGTLTCVATWKSLKAKSATVNLTVIRCPQ
                     SLQNSRLHGSAYLIVQVMGELFIPSVNLVVAAENEPCEVTCLPS:WTWLPDISWELGLL
                     WKLIMWALSDMVVLSVRPMEPIIHVNEPTRSQRDQGGNFTSEMIIHNVEPSDSGNIRC
                     NCCCRCCFCCRRKRGFRIQFOKKSEKEKTNKEFTTESGNENSGYNSDEQKTFDTASLP
                     PKSCESSDPEQRNSSCGPHIORADQRPPRPASHPQASFNLASP:KVSNTTVV"
BASE COUNT      299 a    302 c    288 g    279 t
ORIGIN
```

Query Match   95.5%;   Score 1122.4;   DB 6;   Length 1168;
Best Local Similarity   99.5%;   Pred. No. 9.7e-260;
Matches 1126; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy    351 GCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATCTTGTAGTC 410
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    416 GCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATCTTGTAGTC 475

Qy    411 GCTGAGAAATGAACCTTGTGAAGTTACTTGTTTACCCTCACACTGGACCCCCGGAT     470
          |||||||| |||||||||||||||||||||||||||||||||||||||||| |  |||
Db    476 GCTGAGAATGAACCTTGTGAAGTTACTTGTTTACCCTCACACTGGACCTCCCGGAT     535

Qy    471 ATTTCCTGGGAGCTCGGTCGTCTCCTGGGTCAGCCATTCAAGCTATTATTATTTTGTTCCGGAGCCC 530
          ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    476 ATTTCTGGGAGCTCGGTCGTCTCCTGGGTCAGCCATTCAAGCTATTATTATTTTGTTCCGGAGCCC   530

Qy    471 ACCAATCCTTCTGACGCCGACGTGTACTCTTACAATACGCGTCTGCTGCTGCCCGGTCGT 530
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    536 ATTTCCTGGGAGCTCGGTCGTCTCCTGGGTCAGCCATTCAAGCTATTATTATTTTGTTCC 595

Qy    531 ACGGACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGCCCATTCAAGCTATTATTGGTTTG 590
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    596 ACGGACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGACCCCCACAGAGCAATGGGGACTTTG 655

Qy    591 ACTTGGTGGTGGCTACCTTGGGAAGAGCCTGAAAGGCCCGCAAGTCTGCAACTGTAAATCTCACT 650
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    656 ACTTGGCGTGGCTACCTTGGGAAGAGCCTGAAAGGCCCGCAAGTCTGCAACTGTAAATCTCACT   715

Qy    651 GTGGATTCGGTCGTGTCCCAAGACACTGGAGGTGGTGGTATTAATAATTCCAGGTGTATTATCAAGT 710
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    716 GTGGATTCGGTCGTGTCCCAAGACACTGGAGGTGGTGGTATTAATAATTCCAGGTGTATTATCAAGT   775

Qy    711 TTACCGAGTTTAGGTTTTTTCATTGCCTACTTGGGACTTGGACTAGCAGGC     770
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    776 TTACCGAGTTTAGGTTTTTTCATTGCCTACTTGGGACTTGGACTAGCAGGC     835

Qy    771 ACCAATGCTTCTGACGCCGACGTGTACTCTTACAATACGCGTCTGCTGCTGCCCGGTCGT 830
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    836 ACCAATGCTTCTGACGCCGACGTGTACTCTTACAATACGCGTCTGCTGCTGCCCGGTCGT 895

Qy    831 TGTTGTGTGGCTGCAACTGTCTGCTGCCCGTTGTCTGTGTAGAGAAAAAAGAGGATTT    890
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    896 TGTTGTGTGGCTGCAACTGTCTGCTGCCCGTTGTCTGTGTAGAGAAGAAAGAGGATTT    955

Qy    891 CGTATTCAATTTCAAAGAAGAATCTGAAAAAGAGACAAAAAACTGAACAGTCATGAA    950
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    956 CGTATTCAATTTCAAAGAAGAATCTGAAAAAGAGACAAAAAACTGAACAGTCATCCA    1015

Qy    951 AGTGGAAATGAAAAACTCCGGCTACACAATTCACAATGACAAGACGACACCCTTCT    1010
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1016 AGTGGAAATGAAAAACTCCGGCTACAATTCAGATGACAACATCAGAATCGGCCTGCA  1075

Qy   1011 CTCCCTCCCAAATCCTGTGAATCCGATCCTGACACAAGAACAGTAGCTGTTCATTCCAGTGTTAATCT 1070
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1076 CTCCCTCCCAAATCCTGTGAATCCGATCCTGACACAAGAACAGTAGCTGTTCATTCCAGTGTTAATCT  1135

Qy   1071 CCTCACCAGCGGCGGCTGATCAACGTCCACCAGCCAGCAAGTCACCACAGCCTTCTTTT 1130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1136 CCTCACCAGCGGGGCTGATCAACGTCATCCACAGTCATCCACAGGCTTCTTTT       1195

Qy   1131 AATCTGGCCAGTCCTGAGAAGGTCAGTAATACAAACTGTAGTGTATAG           1175
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1196 AATCTGGCCAGTCCTGAGAAGGTCAGTAATACAAACTCTAGTAGTTTAG          1240
```

```
Qy     44 CCTCCACGGTTCTGGGGTTCTGGTTAATGAAGTGCTAATGAGAAGGCCCCAGAAATGCAAGCAGTCCTT 103
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     37 CCCACCCCGGTTCTGGGGTTCTGGTTAATGAAGTGCTAATGAGAAGGCCCCAAAAATGCAAGCAGTCCTT 96

Qy    104 GAAGGGGCTCCCAGGCGTCCGCTTCAACTGCACCGTCTCCCAGGCGTCGGAAGCTCATCATGTG 163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     97 GAAGGGGCTCCCCAGGCGTCCGCTTCAACTGCACCGTCTCCCAGGCGTCGGAAGCTCATCATGTG 156

Qy    164 GGCTCTTCAGTGACATGTGTGGTTGCTCTAAGGCGTCAGCCCCATGGAGCGCCCATCATCACCA 223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    157 GGCTCTTCAGTGACATGTGGTTGCTCTAAGGCGTCAGCCCCATGGAGCGCCCATCATCACCAATGA 216

Qy    224 CCGCCTTCACCTCTCAGAGGTACGACCAGGCGGCGGGGAACTCACCTGGAGATGATCATCCA 283
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    217 CCGCTCACCTCTCAGAGGTACGACCAGGCGGCGGGGAACTCACCTCGGAGATGATCATCCA   276

Qy    284 CAATGTTGGAGCCCCAGTGATTCGGGGGAACATCAGAATCGAGATCCAGCAGCCAGCTCGGCGCA 343
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    277 CAAATGTTGGAGCCCCAGTGATTCGGGGGAACATCAGAATCGAGATCCAGCAGCCAGCTCGGCGCA  336

Qy    344 TGGATCCTGCTTACCTTACCCTCCAAGTTATGGGGGAGAGCTGTTCATTCCAGTGTTAATCT 403
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    337 TGGATCCTGCTTACCTTACCCTCCAAGTTATGGGGGAGAGCTGTTCATTCCAGTGTTAATCT  396

Qy    404 TGTAGTCGCTGAGAATGAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACCCGGGCT 463
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    397 TGTAGTCGCTGAGAATGAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACCGTGGCT 456

Qy    464 CCCGGATATTCCTGGGAGCTCGGTCGTCTCCTGGGTCAGCCATTCAAGCTATTATTTTGTTCC 523
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    457 CCCGGAATATTCCTGGGAGCTCGGTCGTCTCCTGGGTCAGCCATTCAAGCTATTATTTTGTTCC  516

Qy    524 GGAGCCCGCAGCGACCCTTCAAAAGTGCAGTGACCATCCTGGCTCTGCCCATCAAAATGG 583
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    517 GGAGCCCAGCGACCCTTCAAAAGTGCAGTGACCATCCTGGCTCTGCCCATCAAGCAAATGG 576

Qy    584 GACTTTGACTTGGCGTGGCTACCTTGGAAGAGCCTGAAAGGCCCGCAAGTCTGCAACTGTAA 643
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    577 GACTTTGACTTGGCGTGGCTACCTTGGAAGAGCCTGAAAGGCCCGCAAGTCTGCAACTGTAA  636

Qy    644 TCTCACGTGTGATTCGGTGTCCCAAGACACTGGAGGTGGTATTAATAATTCCAGGTGTATT 703
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    637 TCTCACTGTGATTCGGTGTCCCAAGACACTGGAGGTGGTATTAATAATTCCAGGTGTATT   696

Qy    704 ATCAAGTTACCGAGTTTAGGTTTTTTCATTGCCTACTTGGGACTTGGACT          763
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    697 ATCAAGTTACCGAGTTTAGGTTTTTTCATTGCCTACTTGGGACTTGGACT          756

Qy    764 AGCAGGCACCATGCTTCTGACGCCGACGTGTACTCTTACAATACGCGTCTGCTGCTGCCG 823
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    757 AGCAGGCACCATGCTTCTGACGCCGACGTGTACTCTTACAATACGCGTCTGCTGCTGCCG
```

```
BASE COUNT     580 a    463 c    474 g    534 t
ORIGIN

        /clone="PLACE6017788"
        /tissue_type="placenta"
        /clone_lib="PLACE6"
        /note="cloning vector: pME18SFL3"

Query Match
Best Local Similarity   95.8%;    Score 1125.4;   DB 9;    Length 2051;
Matches 1126; Conservative  0;    Pred. No. 1.9e-260;
                      Mismatches  1;   Indels  0;   Gaps  0;

Qy     49  ACGGTTCTGGGTCTTGGTAATGAAGTCATAGAAGTCAACAGAACACCGTTGAAGG   108
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201  ACGGTTCTGGGTCTTGGTAATGAAGTCATAGAAGTCAACAGAACACCGTTGAAGG   260

Qy    109  GCTCCCAGGCTGCGTTCGCGTTCAACTGCACCGTCTCCCAGGCGGCTGGAAGCTC   168
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261  GCTCCCAGGCTGCGTTCGCGTTCAACTGCACCGTCTCCCAGGCGGCTGGAAGCTC   320

Qy    169  TCAGTGACAATGGTGGTGCTGTAAGCGTCAGCGCCCATCATCACCAATGACCCGCT   228
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321  TCAGTGACAATGGTGGTGCTGTAAGCGTCAGCGCCCATCATCACCAATGACCCGCT   380

Qy    229  TCACCTCTCAGAGATTACGACCAGGCGGGGAAACTTCACCTCGGAGAGTGATCATCCACAATG   288
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381  TCACCTCTCAGAGATTACGACCAGGCGGGGAAACTTCACCTCGGAGAGTGATCATCCACAATG   440

Qy    289  TGGACCCAGTGAATCACAGAATACAGAATGCAGCCTTCACCTTGCCCTGCATGGAT   348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441  TGGACCCAGTGAATCACAGAATACAGAATGCAGCCTTCACCTTGCCCTGCATGGAT   500

Qy    349  CTGCTTACCTTACCGTCCAAGTTATATGGGAGAGCGTGTTCATTCCCAGTTAATCTTTGTAG   408
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    501  CTGCTTACCTTACCGTCCAAGTTATATGGGAGAGCGTGTTCATTCCCAGTTAATCTTTGTAG   560

Qy    409  TCGCTGAGAATGAACCTTGTGAAGTACTTGCTACCCTCACACTGGACCGGCTCCCG   468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    561  TCGCTGAGAATGAACCTTGTGAAGTACTTGCTACCCTCACACTGGACCGGCTCCCG   620

Qy    469  ATATTCCTGGGGAGCTCGGTCTCTGGTCAGCCATTCAAGCTATTATTTTGTTCCGGGAC   528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    621  ATATTCCTGGGGAGCTCGGTCTCTGGTCAGCCATTCAAGCTATTATTTTGTTCCGGGAC   680

Qy    529  CCACGGACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGACCCCACAGAGCAAATGGGACTT   588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    681  CCACGGACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGACCCCACAGAGCAAATGGGACTT   740

Qy    589  TGACTTGGCGTGGCTACCTGGAAGAGAGCCTGAAGGCCCGCAAGTCTGCAACTGTAAATCTCA   648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    741  TGACTTGGCGTGGCTACCTGGAAGAGAGCCTGAAGGCCCGCAAGTCTGCAACTGTAAATCTCA   800

Qy    649  CTGTGATTCGGTGTCCCAAGACAGACAGTGGAGGTGTATTAATATTCCAGTGATTATTATCAA   708
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    801  CTGTGATTCGGTGTCCCAAGACAGACAGTGGAGGTGTATTAATATTCCAGTGATTATTATCAA   860

Qy    709  GTTTACCGAGTTTAGGTTTTTCATTGCCCTACTTGGGCGCAAAGTTGGACTTAGCAG   768
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    861  GTTTACCGAGTTTAGGTTTTTCATTGCCCTACTTGGGCGCAAAGTTGGACTTAGCAG   920

Qy    769  GCACCATGCTTCTGACGCCGACGTGTACTACTTACAAATACGCTGCTGCTCGCCGCCGTC   828
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    921  GCACCATGCTTCTGACGCCGACGTGTACTACTTACAAATACGCTGCTGCTCGCCGCCGTC   980

Qy    829  GTTGTTGTGGTCGTCCCCTTGTCTGTGTTTGCTCTGAGAGAAGAAGAGAT   888
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    981  GTTGTTGTGGTCGTCCCCTTGTCTGTGTTTGCTCTGAGAGAAGAAGAGAT   1040

Qy    889  TTCGTATTCAATTTCAAAAGAAGAAATCTGAAAAAGAAAAATCTGAAGACAG   948
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1041  TTCGTATTCAATTTCAAAAGAAGAAATCTGAAAAAGAAAAATCTGAAGACAG   1100

Qy    949  AAAGTGGAAATGAAAACTCCGGCTACAATTCAGATGAACAAAAGACCAGAGAAACCCGCTT   1008
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1101  AAAGTGGAAATGAAAACTCCGGCTACAATTCAGATGAACAAAAGACCAGAGAAACCCGCTT   1160
```

```
FEATURES          Location/Qualifiers
     source       1..1175
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
     CDS          27..1175
                  /note="unnamed protein product"
                  /codon_start=1
                  /protein_id="CAD28216.1"
                  /db_xref="GI:19575327"
                  /translation="MGLVIFLHGSCSGNEVIEGPQNATVLKGSQARFNCTVSQGWKLI
                  MWALSDMVVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRCSLQN
                  SRLHGSAYLIVQVMGELFIPSVNLVVAVAENEPCEVTCLPSHWTRLPDISWELGLLVSHS
                  SYFVPEPSDLQSAVSILALIPQSNGTLICVATWKSLKARKSATVNLTVICPQDTGG
                  GINIPGVLSSLPLSGFSLPTWGKVGLGLAGTMLLPTCLTTIIRCCCRRRCCGGCNCC
                  RCCFCCRRKRGFRIQFQKSEKEKTNKETEEESGNENSGYNSDEQKTTDFASLPPKSC
                  ESSDPEQRNSSCGPPHQRADQRPPPRASHPQASFNLASPEKVSNTTVV"
BASE COUNT        295 a    307 c    286 g    287 t
ORIGIN

Query Match     100.0%;    Score 1175;   DB 6;   Length 1175;
Best Local Similarity   100.0%;   Pred. No. 2.1e-272;
Matches 1175;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy     1 CTGTCTCGCCATCTGAATAACAAGAAATGGGGCCTTGTGATTTTCCTCCACGGTTCTGGGT 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 CTGTCTCGCCATCTGAATAACAAGAAATGGGGCCTTGTGATTTTCCTCCACGGTTCTGGGT 60

Qy    61 CTGGTAATGAAGTCATAGAAGGCCCCCAGAATGCAACAGTCCTTGAAGGGCTCCCAGGCTC 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 CTGGTAATGAAGTCATAGAAGGCCCCCAGAATGCAACAGTCCTTGAAGGGCTCCCAGGCTC 120

Qy   121 GCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCCCTCTCAGTGACATGG 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 GCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCCCTCTCAGTGACATGG 180

Qy   181 TGGGTGCTRAGCGTGCAGGCCCCATGGAGCCCATGGAGCCTTGTGATTTTCCTCCACGGTTCTCAGA 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TGGGTGCTRAGCGTGCAGGCCCCATGGAGCCCATGGAGCCTTGTGATTTTCCTCCACGGTTCTCAGA 240

Qy   241 GGTACGACGCAGGGGCGGGGAACTTCACCTCGGAGATGATCATACCAACAGTGGGAGCCCAGTG 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GGTACGACGCAGGGGCGGGGAACTTCACCTCGGAGATGATCATACCAACAGTGGGAGCCCAGTG 300

Qy   301 ATTCGGGGAACATCAGATGCACGTCACGCTCTCCAGAACAACAGTCGCCTGCATGGATCGTTGCTTACCTTA 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 ATTCGGGGAACATCAGATGCACGTCACGCTCTCCAGAACAACAGTCGCCTGCATGGATCGTTGCTTACCTTA 360

Qy   361 CCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTTAATCTGTTAGTCGGTGAGAATG 420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTTAATCTGTTAGTCGGTGAGAATG 420

Qy   421 AACCTTGTGAAGTTACTTGTCTTACCCTCACACTGGACCCGGCTCCCGGATATTTCCTGGG 480
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 AACCTTGTGAAGTTACTTGTCTTACCCTCACACTGGACCCGGCTCCCGGATATTTCCTGGG 480

Qy   481 AGCTCGGGTCTTCCTGGTCAGCCCATTCAAGCTATATTATTTTGTTCCGGAGCCCAGACCTTC 540
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 AGCTCGGGTCTTCCTGGTCAGCCCATTCAAGCTATATTATTTTGTTCCGGAGCCCAGACCTTC 540

Qy   541 AAAGTGCAGTGAGCGAGCATCCTGGCCTCTGACCCCACAGAGCAATGGGACTTTGCGTGG 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 AAAGTGCAGTGAGCGAGCATCCTGGCCTCTGACCCCACAGAGCAATGGGACTTTGCGTGG 600

Qy   601 CTACCTGGAAGAGGCCGAAGGCCCCAAGTCGCAACTGTAAATCTCACTGTGATTCGGT 660
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 CTACCTGGAAGAGGCCGAAGGCCCCAAGTCGCAACTGTAAATCTCACTGTGATTCGGT 660

Qy   661 GTCCCCAAGACACTGGAGTGGTTGGTATTAATATTCAAGTTTACCGAGTT 720
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 GTCCCCAAGACACTGGAGTGGTTGGTATTAATATTCAAGTTTACCGAGTT 720

Qy   721 TAGGTTTTTCATTGCCTACTTGGGCAAAAGTTGGACTTGGCACCATGCTTC 780
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 2
AK092516

LOCUS   AK092516   2051 bp   mRNA   linear   PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ35197 fis, clone PLACE6017788, highly similar
            to IGSF5.
ACCESSION   AK092516
VERSION   AK092516.1   GI:21751130
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE6
            clone:PLACE6017788.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,O., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
            and Isogai,T.
  TITLE   NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2051)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE   Direct Submission
  JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.

FEATURES   Location/Qualifiers
     source   1..2051
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"

```
Db   721 TAGGTTTTTCATTGCCTACTTGGGCAAAAGTTGGACTTGGCACCATGCTTC 780

Qy   781 TGACGCCGACGTGACTCTTACAATACGCTTGCCAAAGTTGACTTGGACTTGGACTTGGCTT 840
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 TGACGCCGACGTGACTCTTACAATACGCTTGCCAAAGTTGACTTGGACTTGGACTTGGCTT 840

Qy   841 GCAACTGCCTGCCTGCCGTTGTTGTTCTGCTGTAGAAGAGAAGAGGATTTCGTATTCAAT 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 GCAACTGCCTGCCTGCCGTTGTTGTTCTGCTGTAGAAGAGAAGAGGATTTCGTATTCAAT 900

Qy   901 TTCAAAAGAAATCTGAAAAAGAAGAAGACAAACAAAGAAAAACTGAGACAGAAAAGTGGAAATG 960
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TTCAAAAGAAATCTGAAAAAGAAGAAGACAAACAAAGAAAAACTGAGACAGAAAAGTGGAAATG 960

Qy   961 AAAAACTCCGGGCTACAATTCAGATGAACAAAGAAAAAGACCACAGACAAAAGTGGAAAYG 1020
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 AAAAACTCCGGGCTACAATTCAGATGAACAAAGAAAAAGACCACAGACAAAAGTGGAAAYG 1020

Qy  1021 AATCCTTGTGAATCCAGTGATCCTGAACAAAGAAAAACAGTAGCTGTGGCCCTCCTCACCCAGC 1080
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 AATCCTTGTGAATCCAGTGATCCTGAACAAAGAAAAACAGTAGCTGTGGCCCTCCTCACCCAGC 1080

Qy  1081 GGGCTGATCAACGTCCACCCAGGCCCAGCAGAAGTCATCCACACAGGCTTCTTTTAATTGGCCA 1140
         |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db  1081 GGGCTGATCAACGTCCACCCAGGCCCAGCAGAAGTCATCCACACAGGCTTCTTTTAATTGGCCA 1140

Qy  1141 GTCCTGAGAAGGTCAGTAAATACAACTGTAGTATAG 1175
         ||||||||||||||||||||||||||||||||||||
Db  1141 GTCCTGAGAAGGTCAGTAAATACAACTGTAGTATAG 1175
```

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 23:25:03 ; Search time 3244.94 Seconds
               (without alignments)
               10538.186 Million cell updates/sec

Title: US-09-729-264-1
Perfect score: 1175
Sequence: 1 ctgtctgccactctgaataa..........gtaatacaactgtagtatag 1175

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs. 1455140 2878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :
               GenEmbl:*
               1: gb_ba:*
               2: gb_htg:*
               3: gb_in:*
               4: gb_om:*
               5: gb_ov:*
               6: gb_pat:*
               7: gb_ph:*
               8: gb_pl:*
               9: gb_pr:*
              10: gb_ro:*
              11: gb_sts:*
              12: gb_sy:*
              13: gb_un:*
              14: gb_vi:*
              15: em_ba:*
              16: em_fun:*
              17: em_hum:*
              18: em_in:*
              19: em_mu:*
              20: em_om:*
              21: em_or:*
              22: em_ov:*
              23: em_pat:*
              24: em_ph:*
              25: em_pl:*
              26: em_ro:*
              27: em_sts:*
              28: em_sy:*
              29: em_vi:*
              30: em_un:*
              31: em_htg_hum:*
              32: em_htg_inv:*
              33: em_htg_mus:*
              34: em_htg_other:*
              35: em_htg_pln:*
              36: em_htg_rod:*
              37: em_htg_vrt:*
              38: em_sy:*
              39: em_htgo_hum:*
              40: em_htgo_mus:*
              41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

---

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length DB	ID	Description
1		1175	100.0	1175 6	AX380396	AX380396 Sequence
2		1125.4	95.8	2051 9	AK092516	AK092516 Homo sapi
3		1123.4	95.6	1240 6	AX380400	AX380400 Sequence
4		1122.4	95.5	1168 6	AX380398	AX380398 Sequence
5		1067.4	90.8	1139 6	AX380402	AX380402 Sequence
6		381	32.4	1195 6	AX380404	AX380404 Sequence
7		320.2	27.3	142742 9	AF121782	AF121782 Homo sapl
8		320.2	27.3	340000 9	HS21C080	HS21C080 Homo sapl
9	c	188.6	16.1	196900 6	AC020851	AC020851 Mus muscu
10		186	15.8	754 6	AX380408	AX380408 Sequence
11		186	15.8	895 10	BC004806	BC004806 Mus muscu
12		186	15.8	1556 10	AX380406	AX380406 Sequence
13	c	177.6	15.1	175861 2	AC120346	AC120346 Mus muscu
14		100	8.5	199665 9	AF064857	AF064857 Homo sapl
15	c	58	4.9	175861 2	AC120346	AC120346 Mus muscu
16	c	58	4.9	196900 2	AC020851	AC020851 Mus muscu
17	c	57	4.9	40205 9	AF045450	AF045450 Homo sapl
18	c	57	4.9	170121 9	AF064860	AF064860 Homo sapi
19	c	56.6	4.8	427 9	U63332	U63332 Human super
20	c	55.4	4.7	53908 2	HSU53332	HSU53332 Homo sapi
21	c	55	4.7	1061 2	AC120145	AC120145 Mus muscu
22	c	53	4.5	178052 2	AC118358	AC118358 Rattus no
23	c	52.8	4.5	138885 2	AC017582	AC017582 Drosophil
24	c	52.8	4.5	166249 2	AC093502	AC093502 Drosophil
25	c	52.8	4.5	185404 3	AC104703	AC104703 Drosophil
26	c	52.8	4.5	341319 3	AE003536	AE003536 Drosophil
27	c	51.8	4.4	397 6	BD010479	BD010479 Diagnosti
28	c	51.4	4.4	85584 2	AC099238	AC099238 Rattus no
29	c	50.8	4.3	162971 2	AC103439	AC103439 Rattus no
30	c	50.6	4.3	147185 9	AC020589	AC020589 Homo sapi
31	c	50.6	4.3	165675 2	AC113925	AC113925 Rattus no
32	c	50.6	4.3	195460 2	AC129068	AC129068 Rattus no
33	c	50	4.3	94674 2	AC128466	AC128466 Rattus no
34	c	50	4.3	179494 2	AC113770	AC113770 Rattus no
35	c	50	4.3	185487 2	AC108346	AC108346 Rattus no
36	c	49.8	4.2	299 9	HSA131896	HSA131896 Homo sapi
37	c	49.8	4.2	161170 2	AC129166	AC129166 Rattus no
38	c	49.4	4.2	49845 2	AC100974	AC100974 Mus muscu
39	c	49.4	4.2	160650 3	AC008188	AC008188 Drosophil
40	c	49.4	4.2	160885 2	AC018184	AC018184 Drosophil
41	c	49.4	4.2	162384 2	AC007840	AC007840 Drosophil
42	c	49.4	4.2	172334 2	AC126602	AC126602 Rattus no
43	c	49.4	4.2	173373 3	AC099016	AC099016 Drosophil
44	c	49.4	4.2	306854 3	AE003797	AE003797 Drosophil
45	c	49.4	4.2	312634 2	AC095096	AC095096 Rattus no

ALIGNMENTS

RESULT 1
AX380396
LOCUS   AX380396   1175 bp   DNA   linear   PAT 18-MAR-2002
DEFINITION   Sequence 1 from Patent WO0200710.
ACCESSION   AX380396
VERSION   AX380396.1 GI:19575326
KEYWORDS   .
SOURCE   human.
  ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE   B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 1 03-JAN-2002;

DR WPI; 2000-662424/64.
XX
PT Plasmid vector arraying neuropsychiatric gene more than copy number
PT containing trinucleotide repeats its use thereof -
XX
PS Disclosure; Page 12; 23pp; Korean.
XX
CC This invention relates to the use of a plasmid vector array
CC containing a neuropsychiatric gene containing trinucleotide repeats.
CC Tri-nucleotide repeats and their expansion areknown to be involved
CC in a number of diseases including fragile X syndrome, spinocerebellar
CC ataxia type III, hereditary hypertrophic, cardiomyopathy, dentatorubral
CC and pallidoluysian atrophy (DRPLA), spinocerebellar ataxia, Marfan
CC syndrome; X-linked spinobubar atrophy. The present sequence
CC represents a nucleotide sequence of the invention showing the
CC the CAG tri-nucleotide repeats.
XX
SQ Sequence 403 BP; 121 A; 108 C; 105 G; 69 T; 0 other;

Query Match   4.4%; Score 51.8; DB 21; Length 403;
Best Local Similarity 58.9%; Pred. No. 0.0039;
Matches 89; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 808 GCTGCTGCTGCTGCCGCCGCTCGTTGTTGTGGCTGCAACTCGTCGCTGCCGGTTGTTGTTTCT 867
         |||||||||||| ||| | ||||||| ||||||||||| |||| |||| | ||||||| |
Db 215 GCTGCTGCTGCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156

Qy 868 GCTGGTAGAGAGAAAAAGAGGAGATTTGTATTCAATTTCAAAAGAAAATCTGAAAAGAGAGAGA 927
         |||| | | || | | | |||||||||||||||| |||||||||| ||| |||||||| |
Db 155 GCTGTTGCTCGTCTTTTGCTGCTGTCTGAAAACATTCAAAAGTGAAGTATATTTTAAAAAACAA 96

Qy 928 CAAACAAAGAAAACTGAGACAGAAAGTGGAAA 958
         |||| ||| ||| |||| ||||| || |||
Db 95 AACTTAAAAGAAATAAATACCACCATGAGAAAA 65

RESULT 15
ID ABL16830
AC ABL16830;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 1963.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 1963; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 46954 BP; 13370 A; 9970 C; 9854 G; 13760 T; 0 other;

Query Match   4.2%; Score 49.4; DB 23; Length 46954;
Best Local Similarity 63.6%; Pred. No. 0.022;
Matches 91; Conservative 0; Mismatches 51; Indels 1; Gaps 1;

Qy 771 ACCANGCTTCTGACGCCGACGTGTACTCT-TACAATACCGCTGCTGCTGCCGCCCGTCG 829
         || ||| || ||| | || ||| | | | || | | | || | | ||| ||| |
Db 16836 ACGATGGGTGCCTGGCCCGCCGATGATGTGGATGCTTCTGGATGATGCTGCCTGCAGATGCAA 16895

Qy 830 TTGTTGTTGTGGCTGCAACTCGTCGCTGCCGGTTGTTGTTTCTGCTTAGAAGAAGAAAGAGGATT 889
         |||| || ||| ||| | ||| || |||| | ||||||| ||| | | || || || |||
Db 16896 TTGCGTGCGCCCGCGTGCTGCCTGCGTCGCTGCTGCTGCACTTGCGGATCGGATCGAGTCGA 16955

Qy 890 TCGTATTTCAATTTCAAAAGAAAT 912
         | ||| |||||||||||||| |||
Db 16956 AAAGTTACAATTTCAAAAACAAAT 16978

Search completed: April 29, 2003, 11:04:24
Job time : 344.322 secs

Left column:

```
XX
CC   This invention relates to the use of a plasmid vector array
CC   containing a neuropsychiatric gene containing trinucleotide repeats.
CC   Tri-nucleotide repeats and their expansion are known to be involved
CC   in a number of diseases including fragile X syndrome, spinocerebellar
CC   ataxia type III, hereditary hYpertrophic, cardiomyopathy,
CC   and pallidluysian atrophy (DRPLA), spinocerebellar ataxia, Marfan
CC   syndrome; X-linked spinobubar atrophy. The present sequence represents
CC   a tri-nucleotide repeat related sequence #5 disclosed in the scope
CC   of the invention.
CC   Note: Nucleotides which were illegible in the specification have
CC   been represented by an n to keep nucleotide numbering correct.
XX
SQ   Sequence 389 BP; 107 A; 92 C; 91 G; 93 T; 6 other;

Query Match    4.4%;   Score 51.8;   DB 21;   Length 389;
Best Local Similarity   57.4%;   Conservative   0;   Pred. No. 0.0038;
Matches    89;   Conservative   0;   Mismatches   66;   Indels   0;   Gaps   0;

Qy   808 GCTGCTGCTGCTGCCGCCGCCGTCGTTGTGTGGCCTGCAACTGCTGCTGCCGTTGTTGTTTCT 867
         ||||||||||||||||||||||||||| || |||||||||||| ||||| |||
Db   178 GCTGCTGCCGCTGCTGCCTGCCTCTTGCTGCTGCTGCTGCTGCTGCTGCTGCCGTTGTTGTT 119

Qy   868 GCTGTAGAAGAAAAAGAGGATTTCGTATTCAATTTCAAAAAGAGAAATCTGAAAAAGAGAAGA 927
         |||||  |  | ||| |||| | |||  ||  ||||||| | |||||| |
Db   118 GCTGTTGCTGCTTTTGCTGCTGCTGAAACATTCAAAAAGTGAAGTATATTTAAAAAAACAA 59

Qy   928 CAAACAAAGAAAACTGAGACAGAGAAAGTGGAAAATCAA 962
         ||||  | ||||   | || |||||  ||| |
Db   58 AANTTAAAAGAANAACCTACACCATGAGNTANAAAA 24
```

```
RESULT 13
AAX89891/c
ID   AAX89891 standard; DNA; 397 BP.
XX
AC   AAX89891;
XX
DT   05-NOV-1999  (first entry)
XX
DE   Spinocerebellar ataxia type III (SCAIII) gene fragment.
XX
KW   Spinocerebellar ataxia type III; SCAIII; reverse dot hybridisation;
KW   PCR-microplate hybridisation; PCR-MPH; trinucleotide repeat; TNR;
KW   SCAIII Syndrome; ss.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FT   repeat_region    137..355
FT                    /*tag= a
FT   repeat_unit      137..139
FT                    /*tag= b
FT                    /note= "trinucleotide repeat"
XX
PN   WO9943852-A1.
XX
PD   02-SEP-1999.
XX
PF   18-FEB-1999;   99WO-KR00078.
XX
PR   26-FEB-1998;   98KR-0006278.
XX
PA   (JIND/) JIN D K.
PA   (SMSU ) SAMSUNG FINE CHEM CO LTD.
XX
PI   Jin DK;
XX
DR   WPI; 1999-527634/44.
XX
PT   Diagnosis of spinocerebellar ataxia type III (SCA III) syndrome
PT   using techniques which ensure highly accurate diagnosis
XX
```

Right column:

```
PS   Claim 1; Page 12-13; 28pp; English.
XX
CC   The invention relates to the diagnosis of spinocerebellar ataxia type
CC   III (SCAIII) syndrome using reverse dot hybridisation or PCR-microplate
CC   hybridisation (PCR-MPH). The method comprises attaching a portion of the
CC   SCAIII gene containing 73 copies of the trinucleotide (CAG) repeat unit
CC   (the present sequence) to a substrate, and hybridising with amplified
CC   testee genomic DNA containing copies of the trinucleotide units, PCR
CC   amplified with labeled primers (AAX89989-90). The new method is useful
CC   for diagnosis of SCAIII syndrome, and for determining the severity of
CC   the disease. The present sequence represents the SCAIII gene fragment
CC   containing 73 trinucleotide (TNR) repeats.
XX
SQ   Sequence 397 BP; 124 A; 104 C; 104 G; 65 T; 0 other;

Query Match    4.4%;   Score 51.8;   DB 20;   Length 397;
Best Local Similarity   58.9%;   Pred. No. 0.00039;
Matches    89;   Conservative   0;   Mismatches   62;   Indels   0;   Gaps   0;

Qy   808 GCTGCTGCTGCTGCCGCCGCCGTCGTTGTGTGGCCTGCAACTGCTGCTGCCGTTGTTGTTTCT 867
         ||||||||||||||||||||||||||| || |||||||||||| ||||| |||
Db   200 GCTGCTGCGTGCTGCTGCCTGCCTCTTGCTGCTGCTGCTGCTGCTGCCGTTGTTGTT 141

Qy   868 GCTGTAGAAGAAAAAGAGGATTTCGTATTCAATTTCAAAAAGAAAATCTGAAAAAGAGAAGA 927
         |||||  |  | ||| |||| | |||  ||  ||||||| | |||||| |
Db   140 GCTGTTGCTGCTTTTGCTGCTGCTGAAACATTCAAAAAGTGAAGTATATTTAAAAAAACAA 81

Qy   928 CAAACAAAGAAAACTGAGACAGAGAAAGTGGAAA 958
         ||||  | ||||   | || |||||  |||
Db   80 AACTTAAAAGAATAAAATACACCATGAGANTAAAA 50
```

```
RESULT 14
ABK10240/c
ID   ABK10240 standard; DNA; 403 BP.
XX
AC   ABK10240;
XX
DT   20-MAY-2002  (first entry)
XX
DE   Trinucleotide repeat sequence #3.
XX
KW   Trinucleotide repeat; fragile X syndrome; dsi;
KW   spinocerebellar ataxia type III; DRPLA; neuropsychiatric;
KW   hereditary hypertrophic cardiomyopathy; Marfan syndrome;
KW   dentatorubral and pallidoluysian atrophy;
KW   spinocerebellar ataxia; X-linked spinobubar atrophy.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FT   repeat_region    152..347
FT                    /*tag= a
FT                    /rpt_type= TANDEM
FT   repeat_unit      152..154
FT                    /*tag= b
FT                    /note= "CAG type repeat"
XX
OS   Homo sapiens.
XX
PN   KR2000003004-A.
XX
PD   15-JAN-2000.
XX
PF   25-JUN-1998;   98KR-0024064.
XX
PR   25-JUN-1998;   98KR-0024064.
XX
PA   (SMSU ) SAMSUNG MEDICAL CENT.
PA   (JIND/) JIN D G.
XX
PI   Jin DG;
XX
```

CC   and pallidoluysian atrophy (DRPLA), spinocerebellar ataxia, Marfan
CC   syndrome; X-linked spinobulbar atrophy. The present sequence
CC   represents a nucleotide sequence of the invention showing the
CC   the CAG tri-nucleotide repeats.
XX
SQ   Sequence 277 BP; 80 A; 66 C; 62 G; 69 T; 0 other;

Query Match   4.5%; Score 53.4; DB 21; Length 277;
Best Local Similarity 59.6%; Pred. No. 0.00012;
Matches 90; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

Qy   808 GCTGCTGCTGCTGCCCCGTCGTTGTTGTGGCTGCAACTGCTTGCTCCCGTTGTTGTTTCT 867
         |||||||||||||||||| ||||||||||||||||||| ||||| ||||||||||| |||
Db   215 GCTGCTGCTGCTGCTCGCTGTTGTTGTGGCTGCTGCTGCTGCTCGCTGTTGTTGTGCT 156

Qy   868 GCTGTTAGAAGAAAAAAGAGGATTTCGTATTCAAATTCAAAAAGAAAATCTGAAAAACAGAAGA 927
         |||||||||||||||||||||| |||||||||||||||||||| |||||| |||||||||||||
Db   155 GCTGTTGCTGTTGTTTCGTCGTGTCGAAACATTCAAAAGTGAAGTATATTTAAAAAACAA 96

Qy   928 CAAACANAGAAAACTGAGACAGAAAGTGGAAA 958
         ||| || | |||||||||| |||||| | |||
Db   95 AACTTAAAAGAATAAATACACCATGAGAAAA 65

RESULT 11
ABL29756
ID   ABL29756 standard; DNA; 7029 BP.
XX
AC   ABL29756;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 40741.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 40741; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7029 BP; 1426 A; 1440 C; 1945 G; 2218 T; 0 other;

Query Match   4.5%; Score 52.8; DB 23; Length 7029;
Best Local Similarity 69.2%; Pred. No. 0.00094;
Matches 72; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy   808 GCTGCTGCTGCTGCCCCGTCGTTGTTGTGGCTGCAACTGCTTGCTCCCGTTGTTGTTTCT 867
         |||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||| ||
Db   5500 GCTGCTGCTGCTGCTCGCTGTTGTTGTGGCTGCTGCTGCTGCTCGCTGTTGTTGTGCT 5559

Qy   868 GCTGTTAGAAGAAAAAAGAGGATTTCGTATTGTTGTTGTTGTTGTTGTTGTTGTTGTTGCT 911
         ||||||| |||||||||||||| |||||||||| |||| ||||||||||||| || ||| ||
Db   5560 GCTGTATTTGCTAAAAAGAAGAGTAATCATTAATTCCGGTAAAAA 5603

RESULT 12
ABK10248/c
ID   ABK10248 standard; DNA; 389 BP.
XX
AC   ABK10248;
XX
DT   20-MAY-2002 (first entry)
XX
DE   Tri-nucleotide repeat related sequence #5.
XX
KW   Trinucleotide repeat; fragile X syndrome; ds;
KW   spinocerebellar ataxia type III; DRPLA; neuropsychiatric;
KW   hereditary hypertrophic cardiomyopathy; Marfan syndrome;
KW   dentatorubral and pallidoluysian atrophy;
KW   spinocerebellar ataxia; X-linked spinobubar atrophy.
XX
OS   Unidentified.
XX
FH   Key   Location/Qualifiers
FH
FT   misc_difference 28
FT   /*tag=   a
FT   /note=   "Illegible in the specification"
FT   misc_difference 31
FT   /*tag=   b
FT   /note=   "Illegible in the specification"
FT   misc_difference 46
FT   /*tag=   c
FT   /note=   "Illegible in the specification"
FT   misc_difference 56
FT   /*tag=   d
FT   /note=   "Illegible in the specification"
FT   misc_difference 203
FT   /*tag=   e
FT   /note=   "Illegible in the specification"
FT   misc_difference 276
FT   /*tag=   f
FT   /note=   "Illegible in the specification"
XX
OS   Homo sapiens.
XX
PN   KR200003004-A.
XX
PD   15-JAN-2000.
XX
PF   25-JUN-1998; 98KR-0024064.
XX
PR   25-JUN-1998; 98KR-0024064.
XX
PA   (SMSU) SAMSUNG MEDICAL CENT.
PA   (JIND/) JIN D G.
XX
PI   Jin DG;
XX
DR   WPI; 2000-662424/64.
XX
PT   Plasmid vector arraying neuropsychiatric gene more than copy number
PT   containing trinucleotide repeats its use thereof -
XX
PS   Disclosure; Fig 10; 23pp; Korean.

XX
PF 28-JUN-2001; 2001WO-US20719.
XX
PR 28-JUN-2000; 2000US-214512P;
PR 28-NOV-2000; 2000US-0729264.
XX
PA (AMGE-) AMGEN INC.
XX
PI Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR WPI; 2002-130881/17.
DR P-PSDB; AAU75545.
XX
PT New B7-like polypeptides, polynucleotides and their modulators, useful
PT for diagnosing, preventing and treating reproductive, immune and
PT proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS Claim 1; Fig 6; 135pp; English.
XX
CC The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC The polypeptide, polynucleotide encoding it and antibody against (I) are
CC useful for treating B7-like polypeptide-related disease disorders or
CC conditions including reproductive disorders (e.g. infertility,
CC miscarriage, preterm labour and delivery and endometriosis) and
CC proliferative disorders. Antibodies, soluble proteins comprising
CC extracellular domains and other regulators of B7-L polypeptides are
CC useful for enhancing the immune response to tumours. (I) plays a role in
CC growth and maintenance of cancer cells based on the observation of
CC seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC polypeptide. Hence modulators of (I) are useful for the treatment of
CC cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC in allograft transplantation, graft versus host disease. T-cell
CC dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC molecules are useful for alleviating the symptoms associated with
CC diseases involving chronic immune cell dysfunction or to treat
CC autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC purpura and psoriasis, chronic inflammatory disease such as
CC inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC are also useful as immunosuppressive agents for bone marrow and organ
CC transplantation or to prolong graft survival. B7-L molecules are also
CC useful for diagnosis and treatment of diseases involving abnormal cell
CC proliferation, including arteriosclerosis and vascular restenosis.
CC Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC shock syndrome or allosensitisation due to blood transfusions, and for
CC treatment of allergy, asthma and hypersensitivity reactions,
CC nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC lymphoproliferative disorders such as multiple myeloma. The present
CC sequence represents the coding sequence of mouse B7-L_m2.
XX
SQ Sequence 895 BP; 250 A; 223 C; 212 G; 210 T; 0 other;

Query Match 15.8%; Score 186; DB 24; Length 895;
Best Local Similarity 71.1%;
Matches 246; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 24 GAGATGGGGCTTGTGTGATTTCCTCCACGGTTCTGGGTCTGGTTGGTAATGAAGTCATAGAAGGC 83
        | ||| ||| || | | |||||| ||| ||||| |||| || ||| ||| || | ||| |||
Db 83 GTGCTGGCTCATCCTGGCTCACGGTGACAGCTTCCGGATCCAGTTATCAGATCATAGAAGGT 142

Qy 84 CCCAGAGAATGCAACAGTCCTCGAAGGGCTCCGGCTCGCTTCAACTGCACCGTCTCCCAG 143
        || || || |||| || || |||||| || ||||| || ||| |||||||| ||
Db 143 CCTCAGAATGTAACAGTCCTAAAGGACTCAGAGGCTTCACTACTGCACCGTGACTCAC 202

Qy 144 GGCTGGAAGCTCATCANGTGGGCTCTCAGTGACANGTGGTGTGCTAAGCGTCAGGCCCATG 203
        ||||| | | | ||| | |||| ||||||||| | |||||||| |||||| ||||| |||
Db 203 GGCTGGAAGCTTCTCANGTGGACTCTTCANGTGGTTGGTTGCTCAGTCTCACCACCAA 262

Qy 204 GAGCCCATCATCACCAANTGACCGCGCTTCACCTTCTCCAGAGGTACGGACCAGGGCGGGAACTTC 263
        ||| |||||| ||| ||||| || ||||| || ||| |||||| | ||| |||| ||| ||
Db 263 GGACCCATCATCACCAACAACCGCCTTCACCTTATGCCAGTGACAGCACTGACGCTTC 322

Qy 264 ACCTCGGAGATGATCATCCACAANGTTGGAGCCCCAGTGATTCGGGGAACATCAGATGCAGG 323
        || |||||| || ||| || |||||||||| || ||| || ||||| |||| || ||||
Db 323 ANCTCGGAGTTGATCATCCATCCATGAGTGTCAGCCCCAGTGACGTCTGGGGATCCGTGCAATGCAGG 382

Qy 324 CTCCAGAACAGTCGGCCTCCATGGAGTCGCTTACCTTACCGTGCCAAG 369
        ||||||| || ||||| || | | ||||||| |||| ||||| |||
Db 383 CTGCAGAACAGCCCATGGGTTTGGANTCGCCTCCTCAGTGCAAG 428

XX
RESULT 10
XX
ID ABK10239/c
AC ABK10239;
XX
DT 20-MAY-2002 (first entry)
XX
DE Trinucleotide repeat sequence #2.
XX
KW Trinucleotide repeat; fragile X syndrome; ds; DRPLA;
KW spinocerebellar ataxia type III; Marfan syndrome;
KW hereditary hypertrophic cardiomyopathy; neuropsychiatric;
KW dentatorubral and pallidoluysian atrophy;
KW spinocerebellar ataxia; X-linked spinobubar atrophy.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FH
FT repeat_region 152..194
FT   /rpt_type= a
FT repeat_unit 152..154
FT   /*tag= b
FT   note= TANDEM
FT repeat_region 198..222
FT   /*tag= a
FT   /note= "CAG type repeat"
FT repeat_unit 198..200
FT   /rpt_type= TANDEM
FT   /*tag= b
FT   /note= "CAG type repeat"
XX
OS Homo sapiens.
XX
PN KR200000304-A.
XX
PD 15-JAN-2000.
XX
PF 25-JUN-1998; 98KR-0024064.
XX
PR 25-JUN-1998; 98KR-0024064.
XX
PA (JIND/) JIN D G.
PA (SMSU) SAMSUNG MEDICAL CENT.
XX
PI Jin DG;
XX
DR WPI; 2000-662424/64.
XX
PT Plasmid vector arraying neuropsychiatric gene more than copy number
PT containing trinucleotide repeats its use thereof -
XX
PS Disclosure; Page 12; 23pp; Korean.
XX
CC This invention relates to the use of a plasmid vector array
CC containing a neuropsychiatric gene containing trinucleotide repeats.
CC Tri-nucleotide repeats and their expansion areknown to be involved
CC in a number of diseases including fragile X syndrome, spinocerebellar
CC ataxia type III, hereditary hypertrophic, cardiomyopathy, dentatorubral

RESULT 8
ABK13034
```
ID   ABK13034 standard; cDNA; 754 BP.
XX
AC   ABK13034;
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding mouse B7-like protein, B7-L_m3.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Mus musculus.
XX
FH   Key            Location/Qualifiers
FT   CDS            53..724
FT                  /*tag= a
FT                  /product= "B7-like protein, B7-L_m3"
XX
PN   WO20020200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75546.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 1; Fig 7; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-L polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
```

```
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease) various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of mouse B7-L_m3.
XX
SQ   Sequence 754 BP; 220 A; 191 C; 175 G; 168 T; 0 other;

Query Match    15.8%;  Score 186;  DB 24;  Length 754;
Best Local Similarity   71.1%;  Pred. No. 3.2e-41;
Matches  246;  Conservative   0;  Mismatches 100;  Indels   0;  Gaps   0;

Qy   24 GAGATGGGGGCTTGTTGATTTTCCTCCACGGTTCGTGGTCTGGTAATGAACTCATAGAAGGC 83
         |  ||  ||  ||| || || ||||| | |  || |  |  | | || |  |||||||||||
Db   83 GTGCGTGGTCATCCCTGGCTCAGCCTGACAGCTTTCCGGATCAGATCATAGAAGAGT 142

Qy   84 CCCCAGAATGCCAACAGCTCCTGAAGGCGTCCCAGGCGTTCAACTGCACCGTCTCCCAG 143
         || ||||||| | ||||  |  | ||| || ||||||||||||||||||||||| | |||
Db  143 CCTCAGAATGTAACAGTCCTAAAAGGACTCAGAGGCGTCACTTCAACTGCACCGTGACTCAC 202

Qy  144 GGCTGGAAGCTTCATCATGTGGGCGTTCAGTGACATGGTGGTTGCTAAGGCGCCCATG 203
         |||||||||||| || ||||| | | | || || ||||||||| || |   ||| |||
Db  203 GGCTGGAAGCTTCTCATGTGGACTCTTAACCAAATGGTGGTGGCTGAGTGTCTCACCACCAA 262

Qy  204 GAGCCCATCATCACCATGACCGCGTTCACCTCTCAGAGGTGACCAGGGCGGGGAACTTC 263
         || ||||||||||||| || | |||||||| | | |||  |||| |  |||||| ||||
Db  263 GGACCCATCACCACACCAACACCGGCTTCACCTTATGCCAGTTACACAGCACACTGACAGCTTC 322

Qy  264 ACCTCGGAGATGATCATCCACCAAATGTGGAGCCCCAGTGATTCGGGGAACATCAGATGCAGC 323
         |||||||| ||||||||||| | ||| | || ||| ||||| |||| | |||  ||||| |||
Db  323 ATCTCGGAGTTGATCATCCANGATGTGCAGCCCAGTGACTCGGGATCCGTGCAATGCAC 382

Qy  324 CTCCAGAACAGTCGGCCTGCATGGAGTCTGCTTACCTTTACCGTCCAAG 369
         || |||||  || |  || || |||  |||| | ||| |||||||||||
Db  383 CTGCAGAAACAGCCATGGGTTTGGATCTGCCTTCCTCTCAGTGCAAG 428
```

RESULT 9
ABK13033
```
ID   ABK13033 standard; cDNA; 895 BP.
XX
AC   ABK13033;
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding mouse B7-like protein, B7-L_m2.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Mus musculus.
XX
FH   Key            Location/Qualifiers
FT   CDS            53..865
FT                  /*tag= a
FT                  /product= "B7-like protein, B7-L_m2"
XX
PN   WO20020200710-A2.
XX
PD   03-JAN-2002.
```

RESULT 7
AAI36582
ID   AAI36582 standard; DNA; 401 BP.
XX
AC   AAI36582;
XX
DT   17-OCT-2001 (first entry)
XX
DE   Probe #5268 used to measure gene expression in human placenta sample.
XX
KW   Probe; microarray; human; placenta; antenatal diagnosis;
KW   genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO20015727272-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00663.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207455.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR   WPI; 2001-488897/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human placenta -
CC   The present invention relates to single exon nucleic acid probes (SENP).
CC   The present sequence is one such probe. The probes are useful for
CC   producing a microarray for predicting, measuring and displaying gene
CC   expression in samples derived from human placenta. The probes are useful
CC   for antenatal diagnosis of human genetic disorders.
XX
PS   Claim 25; SEQ ID NO 5268; 654pp; English.
XX
SQ   Sequence 401 BP; 105 A; 98 C; 111 G; 87 T; 0 other;

```
Qy  1164 ACTGTAGTATA 1174
          || ||||| ||
Db  1154 ACTTTAGTGTA 1164
```

Query Match   16.2%; Score 190.4; DB 22; Length 401;
Best Local Similarity   97.0%; Pred. No. 1.4e-42;
Matches 194; Conservative   0; Mismatches   6; Indels   0; Gaps   0;

```
Qy    42 TTCCTCACGGTTCTGGGTCTGGTAATGAAGTCATAGAAGCCCCAGAATGCAACAGTC 101
         | ||| |||||||||||||||||||||||||||||||||| ||||| ||||| ||||
Db   202 TACCTTCCAGGTTCTGGGTCTGGTAATGAAGTCATAGAAGCGCCCCAAAATGCAGAGTC 261

Qy   102 CTGAAAGGGCTCCAGGCGTCGCGTTCAACTGCACGGTCTCCCAGGGCTGGAAGCTCATCATG 161
         ||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   262 CTGAAGGGGCTCCAGGCGTCGCGTTCAACTGCACGGTCTCCCAGGGCTGGAAGCTCATCATG 321

Qy   162 TGGGCTCTCAGTGACATGGGTGGTGCTTAAGCGTCAGGCCGTCAGGCCCATCATCACCAAT 221
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   322 TGGGCTCTCAGTGACATGGGTGGTGCTTAAGCGTCAGGCC... 381

Qy   222 GACGGCTTCACCTCTCAGAG 241
         ||||||||||||||||||||
Db   382 GACGGCTTCACCTCTCAGAG 401
```

```
Qy      301 CGGCTCCCGGGATATTCCTCGGGAGCTCGGTCTCCTGGTTCGCTCAGCCANTTCAAGCTATTATTTT
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 CGGCTCCCGGGATATTCCTCGGGAGCTCGGTCTCCTGGTTCGCTCAGCCANTTCAAGCTATTATTTT 360

Qy      519 GTTCCGGAGCCCACGGCCTTCAAAGTGCAGTGAGCATGTCCTGGCTCTTGACCCCACAGAGC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 GTTCCGGAGCCCACGGCCTTCAAAGTGCAGTGAGCATGTCCTGGCTCTTGACCCCACAGAGC 578

Qy      579 AATGGGACTTTGACTTGGCTGGGCTACCGGAAGGCCCGCAAGTCTGCAACT
            |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 AATGGGACTTTGACTTGGCTGGGCTACCGGAAGGCCCGCAAGTCTGCAACT 480

Qy      639 GTAAATCTCACTGTGATTCGGTGTGTCCCCAAGACACTGGAGTGGTGATTAATAATTCCAGGT
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 GTAAATCTCACTGTGATTCGGTGTGTCCCCAAGACACTGGAGTGGTGATTAATAATTCCAGGT 698

Qy      699 GTATTATCAAGTTTACCGAGTTTAGGTTTTTTTCATTGCCTACTTGGGGCAAAAGTTTGGACTTT
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 GTATTATCAAGTTTACCGAGTTTAGGTTTTTTTCATTGCCTACTTGGGGCAAAAGTTTGGACTTT 600

Qy      759 GGACTAGCAGGCAGGCACCATGCTTCTGACGCGACGGTACTCTTACAATACGGCTGCTGCTGC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GGACTAGCAGGCAGGCACCATGCTTCTGACGCGACGGTACTCTTACAATACGGCTGCTGCTGC 660

Qy      819 TGCCCCGGTGTCGTTGTTGTTGGCTGGCGCAACTGGCCGTTTGTTGTTTGTTTCTGCCTAGGAAGA
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 TGCCCCGGTGTCGTTGTTGTTGGCTGGCGCAACTGGCCGTTTGTTGTTTGTTTCTGCCTAGGAAGA 720

Qy      879 AAAAGAGGATTTCGTATTC----------
            ||||||||||    |||
Db      721 AAAAGAGGTGACATTCTTCTGACATCTTTCCCACATCCCAGAGATTAGGTGCATGCATC 780

Qy      898 -----------
Db      781 AATGGCGTCANTTTGTGTATACTCANTTTATGTCTGCCTTCTTGAGCCTTCACTGTGAGA 840

Qy      898 -------
Db      841 CTGANCCAATACCGTGCACGAAATCTGAAAAGAGAAGAACAAACAACTGAGACAG 900

Qy      949 AAAGTGGAAANGAAAAACTCCGGCTACAAATTCAGATGAACAAAAGACCACAGAACAGCGCTT
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 AAAGTGGAAANGAAAAACTCCGGCTACAAATTCAGATGAACAAAAGACCACAGAACAGCGCTT 960

Qy     1009 CTCTCCCTCCCAAAATCCTGTGAATCCAGTGAATCCAGCGTGATCAAAGAAAAACAGTAGCGTGTGCC
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 CTCTCCCTCCCAAAATCCTGTGAATCCAGTGAATCCAGCGTGATCAAAGAAAAACAGTAGCGTGTGCC 1020

Qy     1069 CTCCTCACCAGCGCGGGCGGGCTGATCAACGT 1094
            ||||||||||||||||||||||||||
Db     1021 CTCCTCACCAGAGAAGACANTAGACCT 1046
```

```
Db      RESULT 6
        ID     ABK13032 standard; cDNA; 1195 BP.
        XX
        AC     ABK13032;
        XX
        DT     23-APR-2002  (first entry)
        XX
        DE     DNA encoding mouse B7-like protein, B7_m1.
        XX
        KW     Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
        KW     antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
        KW     antiinflammatory; dermatological; antipsoriatic; neuroprotective;
        KW     antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
        KW     antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
        KW     reproductive disorder; graft versus host disease; autoimmune disease;
        KW     toxic shock syndrome; allergy; nephropathy; skin disorder;
        KW     endocrinopathy; lymphoproliferative disorder; gene; ss.
        XX
        OS     Mus musculus.
        XX
```

```
FH     Key             Location/Qualifiers
FT     CDS             53..1165
FT                     /*tag=  a
FT                     /product= "B7-like protein, B7-L_m1"
XX
PN     WO20020071O-A2.
XX
PD     03-JAN-2002.
XX
PF     28-JUN-2001; 2001WO-US20719.
XX
PR     28-JUN-2000; 2000US-2145122P.
PR     28-NOV-2000; 2000US-0729264.
XX
PA     (AMGE-) AMGEN INC.
XX
PI     Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR     WPI; 2002-130881/17.
DR     P-PSDB; AAU75544.
XX
PT     New B7-like polypeptides and their modulators, useful
PT     for diagnosing, preventing and treating reproductive, immune and
PT     proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS     Claim 1; Fig 5; 135pp; English.
XX
CC     The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC     The polypeptide, polynucleotide encoding it and antibody against (I)
CC     useful for treating B7-like polypeptide-related disease, disorders or
CC     conditions including reproductive disorders (e.g. infertility,
CC     miscarriage, preterm labour and delivery and endometriosis) and
CC     proliferative disorders. Antibodies, soluble proteins comprising
CC     extracellular domains and other regulators of B7-L polypeptides are
CC     useful for enhancing the immune response to tumours. (I) plays a role in
CC     growth and maintenance of cancer cells based on the observation of
CC     seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC     polypeptide. Hence modulators of (I) are useful for the treatment of
CC     cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC     testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC     pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC     in allograft transplantation, graft versus host disease, T-cell
CC     dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC     molecules are useful for alleviating the symptoms associated with
CC     diseases involving chronic immune cell dysfunction or to treat
CC     autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC     arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC     purpura and psoriasis, chronic inflammatory disease such as
CC     inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC     Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC     are also useful as immunosuppressive agents for bone marrow and organ
CC     transplantation or to prolong graft survival. B7-L molecules are also
CC     useful for diagnosis and treatment of diseases involving abnormal cell
CC     proliferation, including arteriosclerosis and vascular restenosis.
CC     Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC     shock syndrome or allosensitisation due to blood transfusions, and for
CC     treatment of allergy, asthma and hypersensitivity reactions,
CC     nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC     pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC     (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC     thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC     lymphoproliferative disorders such as multiple myeloma. The present
CC     sequence represents the coding sequence of mouse B7-L_m1.
XX
SQ     Sequence 1195 BP; 313 A; 304 C; 291 G; 287 T; 0 other;
```

```
Qy       24 GAGATGGGGGCTTGTGATTTCCTTCCACGGTTCTGGGTCTGGTTAATGAAGTCATTAGAAGGC 83
            ||| ||    |||| |||||| || ||  ||||||  ||||||| ||| ||||| |||||||||
Db       83 GTGCTGGTTCATCCTGGCTCAGCTGACAGCTTCCGGGATCCAGTTATCAGATCATATAGAAGGT 142
```

Query Match   32.4%;   Score 381; DB 24; Length 1195;
Best Local Similarity   61.9%;   Pred. No. 3.4e-95;
Matches   712;   Conservative   0;   Mismatches 370;   Indels 69;   Gaps 4;

RESULT 5
AAS92356

```
ID   AAS92356 standard; cDNA; 1392 BP.
XX
AC   AAS92356;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28160.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28169.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
CC
PS   Claim 1; SEQ ID No 28160; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful for treating
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1392 BP; 338 A; 361 C; 341 G; 352 T; 0 other;
```

```
Query Match             66.3%;  Score 779.2;  DB 23;  Length 1392;
Best Local Similarity   87.3%;
Matches  913;  Conservative    0;  Mismatches   23;  Indels  110;  Gaps  1;
```

```
Qy    159 ATGTGGGCTCTCAGTGACATGGTGGTGCTAAGCGTCAGGCCCATCATCACAC 218
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGTGGGCTCTCAGTGACATGGTGGTGCTAAGCGTCAGGCCCATCATCACAC 60

Qy    219 AATGACCGCTTCACCCTCAGAGGTACGACCAGGGCGGGAACTTCACCTGGGATGATC 278
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AATGACCGCTTCACCCTCAGAGGTACGACCAGGGCGGGAACTTCACCTGGGATGATC 120

Qy    279 ATCCACAATGTGGAGCCCAGTGATTCGGGGAACATCAGATCCAGCCTCCAGAAACAGTCGC 338
          ||||||||||||||||||||||||||||||||||||||||
Db    121 ATCCACAATGTGGAGCCCAGTGATTCGGGGAACATCAGATCCAGCCTCCAGAAACAGTCGC 180

Qy    339 CTGCATGGAATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGAGCTGTTCATTCCCAGTGTT 398
          ||||||||||||||||||||||||||||||||||||||||
Db    181 CTGCATGGAATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGAGCTGTTCATTCCCAGTGTT 240

Qy    399 AATCTTGTAGTCGCTGAGAAATGAAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACC 458
          ||||||||||||||||||||||||||||||||||||||||
Db    241 AATCTTGTAGTCGCTGAGAAATGAAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACC 300

Qy    459 CGGCTCCGGATATTTCCTGGAGCTCGGTCTCCTGGTCTCAGCCATTCAAGCTATTATTTTT 518
```

```
Db    270 CAATGTGGAGCCCAGTGATTCGGGGAACATCAGATCCAGCCTCCAGAAACAGTCGCCTGCA 329
          |||||||||||||||||||||||||||||||||||||||||||||||
Qy    344 TGGGATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGAGCTGTTCATTCCCAGTGTTAATCT 403
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    330 TGGATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGAGCTGTTCATTCCCAGTGTTAATCT 389
Qy    404 TGTAGTCGCTGAGAAATGAAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACCCGGCT 463
          |||||||||||||||||||||||||||||||||||||||
Db    390 TGTAGTCGCTGAGAAATGAAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACCCGGCT 449
Qy    464 CCCGGATATTTCCTGGGAGCTCGGTCTCCTGGTCTCAGCCATTCAAGCTATTATTTTTGTTCC 523
          |||||||||||||||||||||||||||||||||||||||
Db    450 CCCGGATATTTCCTGGGAGCTCGGTCTCCTGGTCTCAGCCATTCAAGCTATTATTTTTGTTCC 509
Qy    524 GGAGCCCAGCGACCTTCAAAGTGCACTGAGCATCTTGGGCTCTCGACCCCACAGAGCGAATGG 583
          |||||||||||||||||||||||||||||||||||||||
Db    510 GGAGCCCAGCGACCTTCAAAGTGCACTGAGCATCTTGGGCTCTCGACCCCACAGAGCGAATGG 569
Qy    584 GACTTTGACTTGCGTGGCTACCTGTGGAAGAGCCTGAAGGCCCGAAGTGTAA 643
          |||||||||||||||||||||||||||||||||||||||
Db    570 GACTTTGACTTGCGTGGCTACCTGTGGAAGAGCCTGAAGGCCCGAAGTGTAA 629
Qy    644 TCTCACTGTGATTCGGTGTCCCCAAGACACTGGAGGTGGTATTAATATTCCAGGTGTGTATT 703
          |||||||||||||||||||||||||||||||||||||||
Db    630 TCTCACTGTGATTCGGTGTCCCCAAGACACTGGAGGTGGTATTAATATTCCAGGTGTGTATT 689
Qy    704 ATCAAGTTACCGAGTTAGGTTTTTCATTGCCTACTTGGGCCAAAGTTGGACTTGGACT 763
          |||||||||||||||||||||||||||||||||||||||
Db    690 ATCAAGTTACCGAGTTAGGTTTTTCATTGCCTACTTGGGCCAAAGTTGGACTTGGACT 749
Qy    764 AGGAGGCACCATGCTTCTGACGCGCGACGTGTACTCGTTGCTGCTGGCTGTGCCG 823
          |||||||||||||||||||||||||||||||||||||||
Db    750 AGGAGGCACCATGCTTCTGACGCGCGACGTGTACTCGTTGCTGCTGGCTGTGCCG 809
Qy    824 CCGTCGTTGTTGTGGCTGCCAACTGCGTCTGCCGTTGTTGTTTCTGCTGTAGAGAAAAG 883
          |||||||||||||||||||||||||||||||||||||||
Db    810 CCGTCGTTGTTGTGGCTGCCAACTGCGTCTGCCGTTGTTGTTTCTGCTGTAGAGAAAAG 869
Qy    884 AGGATTTCGTATTCAATTCAAAAAAAGAGAAATCTGAAAAAGAAGAAACAAACAAAGAAACTGA 943
          ||                       |||||||||||||||||||||||||||||||||||||
Db    870 AG-----------------------GAAATCTGAAAAAGAAGAAACAAACAAAGAAACTGA 907
Qy    944 GACAGAAAGTGGAAATGAAAAACTCCGGCCTACAATTCAGATGAACAAAGAAGACCACAGACAC 1003
          ||||||||||| |||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    908 GACAGAAAGTGAAAATGAAAAACTCCGGCTACAATTCAGATGAACAAAGAAGACCACAGACAC 967
Qy    1004 CGCTCTCTCCCCCCAAATCCTGTGAATCCACAGTGATCCTGAACAAAGAACAGTAGCTG 1063
          |||| |||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    968 CGCTTCTCTCCCCCAAAATCCTGTGAATCCACAGTGATCCTGAACAAAGAACAGTAGCTG 1027
Qy    1064 TGGCCCTCCTCACCCAGCGGGGCTGATCAAACGTCACCAGGCCAGCAAGTCATCCACAGGC 1123
          ||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db    1028 TGGCCCTCCTCACCAGCGGGGCTGATCAAACGTCACCAGGCCAGCAAGTCATCCACAGGC 1087
Qy    1124 TTCTTTAATCTGGCCACAGTCCTGAGAAGGTCAGTAGTAATACAACTGTAGTATA 1174
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    1088 TTCTTTTAATCTGGCCACAGTCCTGAGAAGGTCAGTAGTAATACAACTGTAGTATA 1138
```

```
Qy    644 TCTCACTGTGATTCGGTGTCCCAAGACACTGGAGGTGGTGTATTAATATTCCAGGTGTATT 703
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    637 TCTCACTGTGATTCGGTGTCCCAAGACACTGGAGGTGGTGTATTAATATTCCAGGTGTATT 696

Qy    704 ATCAAGTTTACCGAGTTTAGGTTTTTCATTGCCTACTTGGGGCAAAAGTTGGACTTGGACT 763
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    697 ATCAAGTTTACCGAGTTTAGGTTTTTCATTGCCTACTTGGGGCAAAAGTTGGACTTGGACT 756

Qy    764 AGCAGGCACCATGCTTCTGACGCCGCGACGTGTAGTCTTTACAATACGCTCGTCGTCCTCCCG 823
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    757 AGCAGGCACCATGCTTCTGACGCCGCGACGTGTAGTCTTTACAATACGCTCGTCGTCCTCCCG 816

Qy    824 CCGTGTCGTTGTTGTGGCTGCAACTGCTGCTGCTGCTGTAGGAAGAAAAG 883
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    817 CCGTGTCGTTGTTGTGGCTGCAACTGCTGCTGCTGCTGTAGGAAGAAAAG 876

Qy    884 AGGATTTCGTATTCAATTTCAAAAGAAAATCTGAAAAAGAGAGAAGAAGACAAACAAAGAAAACTGA 943
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    877 AGGATTTCGTATTCAATTTCAAAAGAAAATCTGAAAAAGAGAGAAGAAGACAAACAAAGAAAACTGA 936

Qy    944 GACAGAAAAGTGGAAANTGAAAAACTCCGGCGTTACAATTCAGATGAACAAAAGACACACAGACAC 1003
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    937 GACAGAAAGTGGAAATGAAAAACTCCGGCGTTACAATTCAGATGAACAAAAGAGACAACAGAGACAC 996

Qy   1004 CGCTTCTCTCTCCCTCCCCAAATCCTGTGAATCCCAGTGATCCTGACAACAAAGACAAAGAAGACAC 1063
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    997 CGCTTCTCTCCCTCCCCAAATCCTGTGAATCCCAGTGATCCTGACAACAAAGACAAAGAAGACAC 1056

Qy   1064 TGGGCCTCCTCACCAGGCGGGCGGCTGATCAACGTATCCACCCAGGCCAGCCAAGTCATCCACAGGC 1123
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1057 TGGGCCTCCTCACCAGGCGGGCGGCTGATCAACGTCACCCAGGCCAGCAAGTCATCCACAGGC 1116

Qy   1124 TTCTTTTAATCTGGCCAGTCCTCGAGAAGGTCAGTAATACAACTGTAGTATAG 1175
          ||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1117 TTCTTTTAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG 1168
```

```
RESULT 4
ABK13031
ID   ABK13031 standard; cDNA; 1139 BP.
XX
AC   ABK13031;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h4.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; cytostatic; dermatological; antipsoriatic; antiarthritic;
KW   antidiabetic; haemostatic; antipsoriatic; neuroprotective;
KW   antiasthmatic; nephrotropic; antibacterial; antiulcer; antiallergic;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1134
FT                   /*tag= a
FT                   /product= "B7-like protein, B7-L_h4"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
```

```
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA,  Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis
XX
PS   Claim 1; Fig 4; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h4.
XX
SQ   Sequence 1139 BP; 290 A; 300 C; 283 G; 266 T; 0 other;
```

```
Query Match             90.8%;  Score 1067.4;  DB 24;  Length 1139;
Best Local Similarity   97.5%;
Matches 1103;  Conservative    0;  Mismatches    6;  Indels   22;  Gaps    1;

Qy     44 CCTCCACGGTTCTGGGTGGTGGTAATGAAGTCAATAGAAGGCCCCAGAAATGCAACAGTGT 103
          || |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db     30 CCCACCCGGTTCTGGGTGGTGGTAATGAAGTCAATAGAAGGCCCCAGAAATGCAACAGTGT 89

Qy    104 GAAGGGCGTTCCCAGGCTGCGTCGGCTTCAAACTGCACGTCTCCCAGGGCTGGAAGCTCATGTG 163
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     90 GAAGGGCGTTCCCAGGCTGCGTCGGCTTCAAACTGCACGTCTCCCAGGGCTGGAAGCTCATGTG 149

Qy    164 GGCTCTCAGTGACAATGGGTGGTGCTCAGTGGACTACGACCAGGGCGGGAACTTCACCTGCGATGATCATCACCAATGA 223
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    150 GGCTCTCAGTGACATGGGTGGTGGTGCTAAGCGTCAGGGCCCCATGGAGGCCCATCATCACCAATGA 209

Qy    224 CCGCTTCACCCTCTCAGAGGTACGACCAGGGCGGGAACCTCACCTCGGAGATGATCATCCA 283
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    210 CCGCTTCACCCTCTCAGAGGTACGACCAGGGCGGGAACCTCACCTCGGAGATGATCATCCA 269

Qy    284 CAATGTGGAGGCCCAGTGATTCGGGGRACATCAGATGCAGCCTTCAGAACAGTGCGCTGCA 343
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1016 AGTGGAAATGAAAACTCCGGCTACAATTCAGATGACAAAAGACCACAGAAACCGCTTCT
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1011 CTCCCTCCCAAATCCGTGAATCCAGTGATCCGTGACAACAGTAGCTGTGTGGCCCT 1070

Qy   1016 CTCCCTCCCAAATCCGTGAATCCAGTGATCCGTGACAACAGTAGCTGTGTGGCCCT 1075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1076 CTCCCTCCCAAATCCGTGAATCCAGTGATCCGTGACAACAGTAGCTGTGTGGCCCT 1135

Qy   1071 CCTCACCAGCGGCGGCCTGATCAACGTCACCACCAGGCCACCAAGTCACCTTCTTTTT 1130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1136 CCTCACCAGCGGCGGCCTGATCAACGTCACCACCAGGCCACCAAGTCACCTTCTTTT 1195

Qy   1131 AATCTGGCCAGTCCTGTCTCGAGAAGGTCAGTAATACAACTGTAGTAGTAG 1175
          |||||||||||||||||  |||||||||||||||||||||||||||||||||
Db   1196 AATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTAGTAG 1240
```

RESULT 3
ABK13029

ID   ABK13029 standard; cDNA; 1168 BP.
XX
AC   ABK13029;
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h2.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antiasthmatic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antidiabetic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   8..1168
FT   /*tag= a
FT   /product= "B7-like protein, B7-L_h2"
XX
PN   WO2020007 10-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-2145127.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75541.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 1; Fig 2; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like protein-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h2.
XX
SQ   Sequence 1168 BP; 299 A; 302 C; 288 G; 279 T; 0 other;

Query Match   95.5%; Score 1122.4; DB 24; Length 1168;
Best Local Similarity   99.5%;
Matches 1126; Conservative   0; Mismatches   6; Indels   0; Gaps   0;

Qy   44 CCTCCACGGTTCTGGGTCTGGTGAATGAAGTCATAGAAGGCCCCAGAATGCAACAGTCCT 103
         || ||||||||||||| |||||| ||||||||||| |||| ||||| |||| |||||
Db   37 CCCACCGGTTCTGGGTTCTGGTAATGAAGTCATAGGAAGTCATACAAGAGTCCT 96

Qy   104 GAAGGGCTCCCAGGCTCGCTTCAACTGCCACCGTCTCCCAGGGCTGGAAGTCATCATGTG 163
         |||||| |||||| ||||||||||||||| ||||||||||||||||||||||||||||||
Db   97 GAAGGGGTCCCACGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGTCATCATGTG 156

Qy   164 GGCTCTCAGTGACATGGTGGGTGCTAAGCGTCAGGCCCCATGGAGCCCATCATCACCAATGA 223
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   157 GGCTCTCAGTGACATGGTGGGTGCTAAGCGTCAGGCCCCATGGAGCCCATCATCACCAATGA 216

Qy   224 CCGCTTCACCTCTCAGAGGTACGAACCAGGGGCGGGAACTTCACCTCGGAGATGATCATCCA 283
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   217 CCGCTTCACCTCTCAGAGGTACGAACCAGGGGCGGGAACTTCACCTCGGAGATGATCATCCA 276

Qy   284 CAATGTGGAGCCCAGTCGATTCGGGGAACAATCAGAATCGAGCCTCCAGAACAGTCGCTGCA 343
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   277 CAATGTGGAGCCCAGTCGATTCGGGGAACAATCAGAATCGAGCCTCCAGAACAGTCGCTGCA 336

Qy   344 TGGATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATCT 403
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   337 TGGATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATCT 396

Qy   404 TGTAGTCGCCTGAGAATGAACCTTGTGAAGTTTACTTGTCTACCCTCACACTGGACCCGGCT 463
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   397 TGTAGTCGCCTGAGAATGAACCTTGTGAAGTTTACTTGTCTACCCTCACACTGGACCTGGCT 456

Qy   464 CCCGGATAATTTCCTGGGGAGCCTCGGGTCTCCTGGTCACCCATTCCAACCTATTATTTTGTTCC 523
         ||||||| ||||||||| |||| ||||||||||||||||| ||||||||||||||||||||||||
Db   457 CCCGGAPATTTCCTGGGAGCTCGGGTCTCCTGGTCACCATTCCAAGCTATTATTTTGTTCC 516

Qy   524 GGAGCCCAGCGACGACCTTCAAAGTGCCAGTGAGCATCCTGGCCTCGACCACCAGAGGCAATGG 583
         |||||||||||| |||||||||||||| ||||||||| |||||||||||| ||||||||||||||
Db   517 GGAGCCCAGCGACCTTCAAAGTGCAGTGAGCATGAGATCCTGGCCTCTGACCCCACAGAGCAATGG 576

Qy   584 GACTTTGACTTGCGTGGCTACCTGGAAGAGGCCTGAAGGAGCCCGCAAAGTCTGCAAACTGTAAA 643
         |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   577 GACTTTGACTTGCGTGGCTACCTGGAAGAGGCCTGAAGGAGCCCGCAAGTCTGCAAACTGTAAA 636
```

DT   23-APR-2002 (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h3.
XX
CC   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
CC   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
CC   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
CC   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
CC   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
CC   reproductive disorder; graft versus host disease; autoimmune disease;
CC   toxic shock syndrome; allergy; nephropathy; skin disorder;
CC   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   80..1240
FT   /*tag= a
FT   /product= "B7-like protein, B7-L_h3"
XX
PN   WO20020710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75542.
XX
PS   Claim 1; Fig 3; 135pp; English.
XX
CC   New B7-like polypeptides, polynucleotides and their modulators, useful
CC   for diagnosing, preventing and treating reproductive, immune and
CC   proliferative disorders, e.g. cancer and arteriosclerosis -
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating reproductive disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease. T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions, skin
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h3.
XX
SQ   Sequence 1240 BP; 318 A; 319 C; 305 G; 298 T; 0 other;

Query Match   95.6%; Score 1123.4; DB 24; Length 1240;
Best Local Similarity 99.9%; Pred. No. 4e-301;
Matches 1124; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   51 GGTTCTGGGTCTGGTAATGAAGTGCATAGAAGAAGGCCCCAGAATGCAACAGTCTCTGAAGGGC   110
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   116 GGTTCTGGGTCTGGTAATGAAGTGCATAGAAGAAGGCCCCAGAATGCAACAGTCTCTGAAGGGC   175

Qy   111 TCCCAGGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCTCTC   170
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176 TCCCAGGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCTCTC   235

Qy   171 AGTGACATGGTGGTTGCTAAGCGTGCAGGCCCCATGGAGCCCATCATCACCAAATGACCGCTTC   230
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   236 AGTGACATGGTGGTTGCTAAGCGTGCAGGCCCCATGGAGCCCATCATCACCAAATGACCGCTTC   295

Qy   231 ACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTGGAGATGGATCATCACAAGTGG   290
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   296 ACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTGGAGATGGATCATCACAAGTGG   355

Qy   291 GAGCCCAGTGATTCGGGGAACATCAGATCAGATCCCAGAACAGTCGCCTGACTGGATCT   350
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   356 GAGCCCAGTGATTCGGGGAACATCAGATCAGATCCCAGAACAGTCGCCTGACTGGATCT   415

Qy   351 GCTTACCTTACCGTCCAAGTTATGGGGAGAGCTGTTCATTCCCAGTGTTAAATCTTGTAGTC   410
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   416 GCTTACCTTACCGTCCAAGTTATGGGGAGAGCTGTTCATTCCCAGTGTTAAATCTTGTAGTC   475

Qy   411 GCTGAGAATGAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACCCGTCCCGGAT   470
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   476 GCTGAGAATGAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACCCGTCCCGGAT   535

Qy   471 ATTTCCTGGGAGCTCGGTCTCCCGGTCAGCCAATTCAAGCTAATTATTTTGTTCGGAGCCC   530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   536 ATTTCCTGGGAGCTCGGTCTCCCGGTCAGCCAATTCAAGCTATTATTTTGTTCGGAGCCC   595

Qy   531 AGCGACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGCAGCCCCAGAGCAATGGGACTTTG   590
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   596 AGCGACCTTCAAAGTGCAGTGAGCATCCTGGCTCTGCAGCCCCAGAGCAATGGGACTTTG   655

Qy   591 ACTTGCGTGGGCTACCTGGAAGAGCCCTGAAGGCCCGAAGTCTCCAACTGTAAATCTCACT   650
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   656 ACTTGCGTGGGCTACCTGGAAGAGCCCTGAAGGCCCGAAGTCTCGAACTGTAAATCTCACT   715

Qy   651 GTGATCCGGTGTCCCAGACACTGGAGGTGGTTATTAATTCCAGGGTGTATTTTATCAGT   710
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   716 GTGATTCGGTGTCCCCAAGACACTGGAGGTGGTTATTAATTCCAGGGTGTATTTATCAAGT   775

Qy   711 TTACCGAGTTTAGGTTTTTTCATTGCCTCACTTGGGGCAAAGTTGGACTTGGCAGGC   770
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   776 TTACCGAGTTTAGGTTTTTTCATTGCCTCACTTGGGGCAAAGTTGGACTTGGCAGGC   835

Qy   771 ACCATGCGTTCTGACGCCGCACGTGTACCTGTACTCTTACAATACGGCTGTCGTGCCCGTCGT   830
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   836 ACCATGCGTTCTGACGCCGCACGTGTACCTGTACTCTTACAATACGGCTGTCGTGCCCGTCGT   895

Qy   831 TGTTGTGGGCTGCAACTGCGTCGCCTGCCGTGTCGTTGTTCTCCTGTGAGAAGGAAGATTT   890
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   896 TGTTGTGGGCTGCAACTGCGTCGCCTGCCGTGTCGTTGTTCTCCTGTGAGAAGGAAGATTT   955

Qy   891 CGTATTCAATTTCAAAAAGAAAATCTGAAAAAGAGAAGACAAAACAAAAGAAAACTGAGACAGAGA   950
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   956 CGTATTCAATTTCAAAAAGAAAATCTGAAAAAGAGAAGACAAAACAAAAGAAAACTGAGACAGAGA   1015

Qy   951 AGTGGAAATGAAAAACTCCGGCTACAAATTCAGATGAACAAAAAGACCACAGACACCGCTTCT   1010

```
PR   28-JUN-2000;  2000US-214512P.
PR   28-NOV-2000;  2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI:  2002-130881/17.
DR   P-PSDB; AAU75540.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 1; Fig 1; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis); vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h1.
XX
SQ   Sequence 1175 BP; 295 A; 307 C; 286 G; 287 T; 0 other;

Query Match             100.0%;  Score 1175;  DB 24;  Length 1175;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 1175;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy          1 CTGTCTGCCCATCTGAATAACAGAGATGGGGGTTGTGATTTTCCTCACGGTTCTGGGT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CTGTCTGCCCATCTGAATAACAGAGATGGGGGTTGTGATTTTCCTCACGGTTCTGGGT 60

Qy         61 CTGGTAATGAAAGTCATAGAAGAGCCCCCCAGAAATGCAAACAGTCCTTGAAGGGGCTC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGGTAATGAAAGTCATAGAAGAGCCCCCCAGAAATGCAAACAGTCCTTGAAGGGGCTC 120

Qy        121 GCTTCAAACTGCACCCTCTCCCAGGCGTGGAAGCTCATCATGTGGGCGCCCAGGTG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GCTTCAAACTGCACCCTCTCCCAGGCGTGGAAGCTCATCATGTGGGCGCCCAGGTG 180

Qy        181 TGGTGCTAAGCGTCAGCCCCATGATGAGCCCATCATCACCAATGACCGGCTTCACCTCTCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGTGCTAAGCGTCAGCCCCATGATGAGCCCATCATCACCAATGACCGGCTTCACCTCTCAGA 240

Qy        241 GGTACGACCAGGCGGGCGGGGAACTTCACCTCGGGAGATGATCATCACCAAATGTGGAGCCCAGTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GGTACGACCAGGCGGGCGGGGAACTTCACCTCGGGAGATGATCATCACCAAATGTGGAGCCCAGTG 300

Qy        301 ATTCGGGGAACATCAGATGCAGCCTCCAGAAACAGTGCGCATGGGATCTGCTTACCTTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 ATTCGGGGAACATCAGATGCAGCCTCCAGAAACAGTGCGCATGGGATCTGCTTACCTTA 360

Qy        361 CCGTCCAAGTTATGGGAGAGCTGGTTCATTCCCAGTGTTAATCTTGTGTGTCGCTGAAATG 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCGTCCAAGTTATGGGAGAGCTGGTTCATTCCCAGTGTTAATCTTGTGTGTCGCTGAAATG 420

Qy        421 AACCTTGTGAAGTTACTTGTGTCTACCCTCACACTGGACCCGGCCTCCCGGGATATTTCCTGGG 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AACCTTGTGAAGTTACTTGTGTCTACCCTCACACTGGACCCGGCCTCCCGGGATATTTCCTGGG 480

Qy        481 AGCTCGGTCTCCTGGTCAGCCATTCAAGCTATATTATTTGTTCCGGAGCCCAGGCGGACCTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AGCTCGGTCTCCTGGTCAGCCATTCAAGCTATATTATTTGTTCCGGAGCCCAGGCGGACCTTC 540

Qy        541 AAAGTGCAGTGAGCATCCTGGCTCTGACCCCACAGCCAATGGGGACTTTGACTTGGCTGG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AAAGTGCAGTGAGCATCCTGGCTCTGACCCCACAGCCAATGGGGACTTTGACTTGGCTGG 600

Qy        601 CTACCTGGAAGAGCCCGAAGTGCTGTCGCACAGAGCCAATGGGACTTTGACTTGGCTG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CTACCTGGAAGAGCCCGAAGTGCTGTCGCACAGAGCCAATGGGACTTTGACTTGGCTG 660

Qy        661 GTCCCCAAGACACTGGAGGTGGTATTAATATTCCAGTGTATTATTATCAAGTTTACCGAGTT 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GTCCCCAAGACACTGGAGGTGGTATTAATATTCCAGTGTATTATTATCAAGTTTACCGAGTT 720

Qy        721 TAGGTTTTTCATTGCCTACTTGGGCAAAGTTGGACTTAGCAGGCACCAGCTTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TAGGTTTTTCATTGCCTACTTGGGCAAAGTTGGACTTAGCAGGCACCAGCTTC 780

Qy        781 TGACGCCGACGTGTACTCTTACAATACGGTGCTGCTGCCGTGCGTGTTGTTGGGCT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TGACGCCGACGTGTACTCTTACAATACGGTGCTGCTGCCGTGCGTGTTGTTGGGCT 840

Qy        841 GCAACTGCTGCTGCGGCCGGTTGTGTTTCTGTGAGAGAGAAAAAGAGAGATTTCGTAT CAAT 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCAACTGCTGCTGCGGCCGGTTGTGTTTCTGTGAGAGAGAAAAAGAGAGATTTCGTAT CAAT 900

Qy        901 TTCAAAAGAAATCTGAAAAGAGAGAAGACAAACAAAGAAAGGAAGAAAGTGGAAATG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCAAAAGAAATCTGAAAAGAGAGAAGACAAACAAAGAAAGGAAGAAAGTGGAAATG 960

Qy        961 AAAACTCCGGCTACAATTCAGATGAACAAAGACCACAGAAGCGTTCTCCCTCCA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAAACTCCGGCTACAATTCAGATGAACAAAGACCACAGAAGCGTTCTCCCTCCA 1020

Qy       1021 AATCCTGTGAATCCAGTGATCTGAACAAAAGAAAACAGTAGTGTGGCCTCCTCACCAGC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AATCCTGTGAATCCAGTGATCTGAACAAAAGAAAACAGTAGTGTGGCCTCCTCACCAGC 1080

Qy       1081 GGGCTGATCAACGTCCACCCAGGCCAGCCAAGTCACCACAGGCTTCTTTTTAATCTGGCCA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGGCTGATCAACGTCCACCCAGGCCAGCCAAGTCACCACAGGCTTCTTTTTAATCTGGCCA 1140

Qy       1141 GTCCTGAGAAGGTCAGTAATACAAACTGTAGTATTAG 1175
              |||||||||||||||||||||||||||||||||||||
Db       1141 GTCCTGAGAAGGTCAGTAATACAAACTGTAGTATTAG 1175

XX
RESULT 2
ABKI3030
XX
ID   ABKI3030 standard; cDNA; 1240 BP.
XX
AC   ABKI3030;
XX
```

```
OM nucleic - nucleic search, using sw model

                    GenCore version 5.1.4_p5_4578
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:             April 28, 2003, 23:25:03 ; Search time 284.322 Seconds
                                              (without alignments)
                                              9306.701 Million cell updates/sec

Title:              US-09-729-264-1
Perfect score:      1175
Sequence:           1 ctgtctgccatctgaataa..........gtaatacaactgtagtatag 1175

Scoring table:      IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0

Searched:           2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters:     4370478

Post-processing:    Minimum Match 0%
                    Maximum Match 100%
                    Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :          1: N_Geneseq_101002:*
                    2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                    3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                    4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                    5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                    6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                    7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                    8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                    9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                   10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                   11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                   12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                   13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                   14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                   15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                   16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                   17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                   18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                   19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                   20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                   21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                   22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                   23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                   24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result                  %      Query
No.     Score   Match  Length  DB  ID          Description
-----   -----   -----  ------  --  --          -----------
 1      1175    100.0   1175   24  ABKI3028    DNA encoding human
 2      1123.4   95.6   1240   24  ABKI3030    DNA encoding human
 3      1122.4   95.5   1168   24  ABKI3029    DNA encoding human
 4      1067.4   90.8   1139   24  ABKI3031    DNA encoding human
 5      779.2    66.3   1392   23  AAS92356    DNA encoding novel
 6      381      32.4   1195   24  ABKI3032    DNA encoding mouse
 7      190.4    16.2    401   22  AAI36582    Probe #5268 used t
 8      186      15.8    754   24  ABKI3034    DNA encoding mouse
 9      186      15.8    895   24  ABKI3033    DNA encoding mouse
```

```
 c  10   53.4    4.5    277   21  ABKI0239    Trinucleotide repe
 c  11   52.8    4.5   7029   23  ABL29756    Drosophila melanog
 c  12   51.8    4.4    389   21  ABKI0248    Tri-nucleotide rep
 c  13   51.8    4.4    397   20  AAX89891    Spinocerebellar at
 c  14   51.8    4.4    403   21  ABKI0240    Trinucleotide repe
 c  15   49.4    4.2  46954   23  ABLI6830    Drosophila melanog
 c  16   48.4    4.1   3217   10  AAN11578    Rat androgen recep
 c  17   48.4    4.1   3217   12  AAQ12002    Full-length rat an
 c  18   48.4    4.1   4180   10  AAN11773    Rat androgen recep
 c  19   47.8    4.1    381   23  ABV44466    Human prostate exp
 c  20   47.6    4.1   2733   20  AAX84442    Mouse brain CNG-1
 c  21   47.4    4.0   3894   23  ABL03353    Drosophila melanog
 c  22   47.4    4.0  22341   23  ABL03352    Drosophila melanog
 c  23   46.6    4.0   2333   22  AAF75339    Human TGF-beta rec
 c  24   46.4    3.9   1983   23  ABL24483    Drosophila melanog
 c  25   46.4    3.9   4044   23  ABL24482    Drosophila melanog
 c  26   45.8    3.9   2682   23  ABLI6097    Drosophila melanog
 c  27   45.8    3.9   8821   23  ABLI6096    Drosophila melanog
 c  28   45.6    3.9   1509   23  ABL24571    Drosophila melanog
 c  29   45.6    3.9   3884   23  ABL24570    Drosophila melanog
 c  30   45.4    3.9   3042   23  ABL28445    Drosophila melanog
 c  31   45.4    3.9   3135   23  ABL25653    Drosophila melanog
 c  32   45.4    3.9   5215   23  ABL28444    Drosophila melanog
 c  33   45.4    3.9   5447   23  ABL25652    Drosophila melanog
 c  34   45.2    3.9   5397   23  ABL29757    Drosophila melanog
 c  35   45      3.8    462   23  ABL26941    Drosophila melanog
 c  36   45      3.8   2115   23  ABL26940    Drosophila melanog
 c  37   45      3.8   2260   23  ABL03191    Drosophila melanog
 c  38   45      3.8   3210   23  ABL08977    Drosophila melanog
 c  39   45      3.8   3624   23  ABL03190    Drosophila melanog
 c  40   45      3.8   5975   23  ABL08976    Drosophila melanog
 c  41   45      3.8   6541   23  ABL04172    Drosophila melanog
 c  42   44.8    3.8    254   21  AAZ45758    Polymorphic region
 c  43   44.8    3.8   1310   20  AAV08558    Transgene for epit
 c  44   44.8    3.8   2614   24  ABK36100    cDNA sequence #491
 c  45   44.8    3.8   3263   20  AAV08560    Transgene for epit
```

ALIGNMENTS

```
RESULT 1
ABKI3028
ID   ABKI3028 standard; cDNA; 1175 BP.
XX
AC   ABKI3028;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L.h1.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             27..1175
FT                   /*tag= a
FT                   /product= "B7-like protein, B7-L_h1"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
```

CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/185,432
    FILING DATE: 21-JAN-1994
    CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
    NAME: Misrock, S. Leslie
    REGISTRATION NUMBER: 18,872
    REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (212) 790-9090
    TELEFAX: (212) 869-8864/9741
    TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
    LENGTH: 5063 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: unknown
MOLECULE TYPE: cDNA/DNA (genomic)
FEATURE:
    NAME/KEY: CDS
    LOCATION: 864..1349
FEATURE:
    NAME/KEY: CDS
    LOCATION: 1944..3370
FEATURE:
    NAME/KEY: CDS
    LOCATION: 3439..3736

US-08-185-432-1

Query Match   3.5%;   Score 41; DB 1; Length 5063;
Best Local Similarity 66.3%;   Pred. No. 0.11;
Matches   59;   Conservative   0; Mismatches   30; Indels   0; Gaps   0;

Qy   808 GCTGCTGCTGCTGCCGCCGTCGTTGTTGTGGGCTGCAACTGCTGCTGCCGTTGTTGTTTCT 867
           |||||||||||||||| || |||||| || || ||||| ||| || |||||| || || ||
Db   2988 GCTGCCTGCTGCCGCCTGTTGCCTGTTGCCATCTGCAGCTGTTGCCTGCTGCTTGTTGCTGCT 2929

Qy   868 GCTGTAGAAGAAAAAGAGGATTTCGTATT 896
           ||||| ||| || || ||| ||||
Db   2928 GCTGTTGCTGTCGCGGATGCAGTGGTAAT 2900

RESULT 15
US-08-479-913E-1/c
Sequence 1, Application US/08479913E
Patent No. 6416998
GENERAL INFORMATION:
    APPLICANT: O'Malley, Bert W.
    APPLICANT: Tsai, Ming-Jer
    APPLICANT: Ledebur, Harry C.
    APPLICANT: Kittle, Joseph D.
    TITLE OF INVENTION: MODIFIED STEROID HORMONES FOR GENE THERAPY AND METHODS FOR THEIR
    FILE REFERENCE: 212/133
    CURRENT APPLICATION NUMBER: US/08/479,913E
    CURRENT FILING DATE: 1995-06-07
    PRIOR APPLICATION NUMBER: US 07/939,246
    PRIOR FILING DATE: 1992-09-02
    NUMBER OF SEQ ID NOS: 6
    SOFTWARE: Microsoft Word 97
SEQ ID NO 1
    LENGTH: 6177
    TYPE: DNA
    ORGANISM: Artificial sequence
    FEATURE:
    NAME/KEY: misc_feature
    OTHER INFORMATION: Plasmid pGRO403R
US-08-479-913E-1

Query Match   3.5%;   Score 40.8; DB 4; Length 6177;
Best Local Similarity 75.0%;   Pred. No. 0.14;
Matches   51;   Conservative   0; Mismatches   17; Indels   0; Gaps   0;

Search completed: April 29, 2003, 10:49:16
Job time : 88.1906 secs

Qy   808 GCTGCTGCTGCTGCCGCCGTCGTTGTTGTGGGCTGCAACTGCTGCTGCCGTTGTTGTTTCT 867
           |||||||||||||| || || ||| || | |||||| ||| || || |||||| || || ||
Db   4785 GCTGCTGCTGCTGCCTGCTGGCTGCTGCTGCTCGCTGCTGCTCGTGCTGCTGCTGCTGCGGCT 4726

Qy   868 GCTGTAGA 875
           |||| | |
Db   4725 GCTGCACA 4718

RESULT 12
US-08-185-432-3/c
Sequence 3, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
  APPLICANT: Artavanis-Tsakonas, Spyridon
  APPLICANT: Busseau, Isabelle
  APPLICANT: Diederich, Robert J.
  APPLICANT: Xu, Tian
  APPLICANT: Matsuno, Kenji
  TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
  TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
  NUMBER OF SEQUENCES: 23
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: PENNIE & EDMONDS
    STREET: 1155 Avenue of the Americas
    CITY: New York
    STATE: New York
    COUNTRY: U.S.A.
    ZIP: 10036-2711
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/185,432
    FILING DATE: 21-JAN-1994
    CLASSIFICATION: 530
  ATTORNEY/AGENT INFORMATION:
    NAME: Misrock, S. Leslie
    REGISTRATION NUMBER: 18,872
    REFERENCE/DOCKET NUMBER: 7326-006
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (212) 790-9090
    TELEFAX: (212) 869-8864/9741
    TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 3771 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: unknown
  MOLECULE TYPE: cDNA
  FEATURE:
    NAME/KEY: CDS
    LOCATION: 345..2558
US-08-185-432-3

Query Match   3.5%;   Score 41;   DB 1;   Length 3771;
Best Local Similarity   66.3%;
Matches   59; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy   808 GCTGCTGCTGCTCCGGCCGCGTGCTGGCGTCCG 867
         ||||||||||| | | ||||| | |
Db 1875 GCTGCTGCTGCCGCTGTTGCTGTTGCATCTCAGCTGTTGCTGCTGCT 1816

Qy   868 GCTGTAGAAGAAAAAGAGGAGATTCGTATT 896
         |||| | | ||| | | |||
Db 1815 GCTGTTGCTGCTCGCGGATGCAGTGGTAAT 1787

RESULT 13
US-08-185-432-23
Sequence 23, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
  APPLICANT: Artavanis-Tsakonas, Spyridon
  APPLICANT: Busseau, Isabelle
  APPLICANT: Diederich, Robert J.
  APPLICANT: Xu, Tian
  APPLICANT: Matsuno, Kenji
  TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
  TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
  NUMBER OF SEQUENCES: 23
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: PENNIE & EDMONDS
    STREET: 1155 Avenue of the Americas
    CITY: New York
    STATE: New York
    COUNTRY: U.S.A.
    ZIP: 10036-2711
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/185,432
    FILING DATE: 21-JAN-1994
    CLASSIFICATION: 530
  ATTORNEY/AGENT INFORMATION:
    NAME: Misrock, S. Leslie
    REGISTRATION NUMBER: 18,872
    REFERENCE/DOCKET NUMBER: 7326-006
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (212) 790-9090
    TELEFAX: (212) 869-8864/9741
    TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 23:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 3771 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: double
    TOPOLOGY: unknown
  MOLECULE TYPE: cDNA
US-08-185-432-23

Query Match   3.5%;   Score 41;   DB 1;   Length 3771;
Best Local Similarity   66.3%;
Matches   59; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

Qy   808 GCTGCTGCTGCTCCGGCCGCGTGCTGGCGTCCGTTCT 867
         ||||||||||| | | ||||| | ||||
Db 1897 GCTGCTGCTGCCGCTGTTGCTGTTGCATCTCAGCTGTTGCTGCTGCT 1956

Qy   868 GCTGTAGAAGAAAAAGAGGAGATTCGTATT 896
         |||| | | ||| | | |||
Db 1957 GCTGTTGCTGCTCGCGGATGCAGTGGTAAT 1985

RESULT 14
US-08-185-432-1/c
Sequence 1, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
  APPLICANT: Artavanis-Tsakonas, Spyridon
  APPLICANT: Busseau, Isabelle
  APPLICANT: Diederich, Robert J.
  APPLICANT: Xu, Tian
  APPLICANT: Matsuno, Kenji
  TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
  TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
  NUMBER OF SEQUENCES: 23
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: PENNIE & EDMONDS
    STREET: 1155 Avenue of the Americas
    CITY: New York
    STATE: New York
    COUNTRY: U.S.A.
    ZIP: 10036-2711
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.30

PCT-US92-11107-12

Query Match 3.6%; Score 42.6; DB 5; Length 3292;
Best Local Similarity 64.9%; Pred. No. 0.031;
Matches 63; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy  804 ATACGCGTGCTGCTGCTGCCGCCGCCGCGCGTCTGTTGTTGGCTGCAACTGTCTGCCGTTGTTGTTGT  863
            | |||| |||| | |||| ||||| ||||| | |||||||||||||| || |||| ||||||||||
Db  818 AGATGATGAATGTGTTGTTGTTGTTGAAGTTGTTGTTGTTGTTGCAACTGGTGTTTGCTGTTGTTGTTGTAGC  759

Qy  864 TTCTGCTGTGAGAAGAAAAAGAGGATTTCGTATTCAAT  900
            |||||||| |  || || ||| | ||||| ||| |||
Db  758 TGCTGCTGTTGCTGCAACTGCTGTTGTTGTGTAATT  722
```

RESULT 9
US-09-043-303-12/c
Sequence 12, Application US/09043303
Patent No. 6251589
GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 78
; TYPE: DNA
; ORGANISM: Homo sapiens

Query Match 3.5%; Score 41.6; DB 4; Length 78;
Best Local Similarity 78.1%; Pred. No. 0.0085;
Matches 50; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```
Qy  808 GCTGCTGCTGCTGCCGCGCGGTCGTTGTTGTGGCGTCGCTGCAACTGCTGTTGTGGCTGCAACTGTTTCT  867
        ||||||||||||| ||| || | |||||| |||  | |||||| || || ||||| ||||||
Db   67 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  867

Qy  868 GCTG  871
        ||||
Db    7 GGTG  4
```

Best Local Similarity 78.1%; Pred. No. 0.014; Mismatches 14; Indels 0; Gaps 0;
Matches 50; Conservative 0;

```
Qy  808 GCTGCTGCTGCTGCCGCGCGGTCGTTGTTGTGGCGTCGCTGCAACTGCTGTTGTGGCTGCAACTGTTTCT  867
        ||||||||||||| ||| || | |||||| |||  | |||||| || || ||||| ||||||
Db   67 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  867

Qy  868 GCTG  871
        ||||
Db    7 GGTG  4
```

RESULT 11
US-08-145-617-5/c
Sequence 5, Application US/08145617
Patent No. 5766847
GENERAL INFORMATION:
; APPLICANT: Jackle, Herbert
; APPLICANT: Tautz, Diethard
; TITLE OF INVENTION: PROCESS FOR ANALYZING LENGTH
; TITLE OF INVENTION: POLYMORPHISMS IN DNA REGIONS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
; STREET: 301 N. Washington Street, P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: United States of America
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/145,617
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/681,494
; FILING DATE: 10-JUN-1991
; APPLICATION NUMBER: DE P3834636.2
; FILING DATE: 11-OCT-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Svensson, Leonard R.
; REGISTRATION NUMBER: 30,330
; REFERENCE/DOCKET NUMBER: 147-122PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-241-1300
; TELEFAX: 703-241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 379 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-145-617-5

Query Match 3.5%; Score 41.2; DB 1; Length 379;
Best Local Similarity 67.4%; Pred. No. 0.025;
Matches 58; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

```
Qy  786 CCGACGTGTACTCTTACAATACCGCTGCGTGCTGCCGGCGTCGTTGTTGTGGGCTGCGCAAC  845
        |||| | |  ||| | | |||| |  |  | | || |||| | |||||  ||| ||| ||
Db  214 CCGAACTCCAGGCCTCCAGTTGCTCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGT  155

Qy  846 TGCTGCTGCCGCGTTGTTGTTTCTGCTG  871
        |||||| |||| | |||| |||| |||||
Db  154 TGCTGCTGCTGATGTTGCTGCTGCTGTTG  129
```

RESULT 10
US-09-043-303-7/c
Sequence 7, Application US/09043303
Patent No. 6251589
GENERAL INFORMATION:
; APPLICANT: TSUJI, Shoji
; APPLICANT: SANPEI, Kazujiro
; TITLE OF INVENTION: Method for Diagnosing Spinocerebellar Alaxia Type 2 and
; TITLE OF INVENTION: Primers Therefor
; FILE REFERENCE: 0760-0241P
; CURRENT APPLICATION NUMBER: US/09/043,303
; CURRENT FILING DATE: 1998-05-18
; EARLIER APPLICATION NUMBER: PCT/JP96/01999
; EARLIER FILING DATE: 1996-07-18
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 203
; TYPE: DNA
; ORGANISM: p-2093 plasmid
US-09-043-303-7

Query Match 3.5%; Score 41.6; DB 4; Length 203;

```
            MOLECULE TYPE: DNA (genomic)
            ORIGINAL SOURCE:
              ORGANISM: Saccharomyces cerevisiae
            IMMEDIATE SOURCE:
              CLONE: lambda yPt
            FEATURE:
              NAME/KEY: CDS
              LOCATION: 1..1626
            US-08-258-442-12

Qy  804  ATACGCTGCTGCTGCTGCGCCGCGCCGCCGGTCGTTGGTTGTTGGCTGCCAACTGCTCGCTTGCCGTTGTTGT  863
         | |||| |   |||  ||  ||    |||  | ||| ||| |  ||| ||||| |  |  ||| |  || ||||
Db  818  AGATGATGATGTGTTGTTCGTTGAAGTTGTTGTTGTTGTTGCAACTGGTGTTCCTGTTGTTGTAGC  759

Qy  864  TTCTGCTGTAGAAGAAAAAGAGGATTTCGTATTCAAT  900
         | |||| || | |  |  |  || |||| | ||| ||
Db  758  TGCTGCTGTTGCTGCAACTGCTGTTGTTGTTGTTAATT  722


RESULT 7
US-08-328-809-7/c
Sequence 7, Application US/08328809
Patent No. 5705334
GENERAL INFORMATION:
  APPLICANT: Lippard, Stephen J.
  APPLICANT: Essigmann, John M.
  APPLICANT: Donahue, Brian A.
  APPLICANT: Toney, Jeffrey H.
  APPLICANT: Bruhn, Suzanne L.
  APPLICANT: Pil, Pieter M.
  APPLICANT: Brown, Steven
  APPLICANT: Kellett, Patti
  TITLE OF INVENTION: Uses For DNA Structure-Specific
  TITLE OF INVENTION: Recognition Proteins
  NUMBER OF SEQUENCES: 8
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Patent Administrator, Testa, Hurwitz & Thibeault
    STREET: 53 State Street
    CITY: Boston
    STATE: MA
    COUNTRY: USA
    ZIP: 02109
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/328,809
    FILING DATE:
  CLASSIFICATION: 435
  ATTORNEY/AGENT INFORMATION:
    NAME: Fenton, Gillian M.
    REGISTRATION NUMBER: 36,508
    REFERENCE/DOCKET NUMBER: MIT-023 (5473/24)
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 617-248-7000
    TELEFAX: 617-248-7100
  INFORMATION FOR SEQ ID NO: 7:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 3292 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: double
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    ORIGINAL SOURCE:
      ORGANISM: Saccharomyces cerevisiae
    IMMEDIATE SOURCE:
      CLONE: lambda yPt
```

```
            FEATURE:
              NAME/KEY: CDS
              LOCATION: 1..1626
            US-08-328-809-7

Query Match   Best Local Similarity 64.9%;  Score 42.6;  Pred. No. 0.031;  DB 1;  Length 3292;
Matches  63;  Conservative  0;  Mismatches  34;  Indels  0;  Gaps  0;

Qy  804  ATACGCTGCTGCTGCTGCGCCGCGCCGCCGGTCGTTGGTTGTTGGCTGCCAACTGCTCGCTTGCCGTTGTTGT  863
         | |||| |   |||  ||  ||    |||  | ||| ||| |  ||| ||||| |  |  ||| |  || ||||
Db  818  AGATGATGATGTGTTGTTCGTTGAAGTTGTTGTTGTTGTTGCAACTGGTGTTCCTGTTGTTGTAGC  759

Qy  864  TTCTGCTGTAGAAGAAAAAGAGGATTTCGTATTCAAT  900
         | |||| || | |  |  |  || |||| | ||| ||
Db  758  TGCTGCTGTTGCTGCAACTGCTGTTGTTGTTGTTAATT  722


RESULT 8
PCT-US92-11107-12/c
Sequence 12, Application PCT/US9211107
GENERAL INFORMATION:
  APPLICANT: Donahue, Brian A.
  APPLICANT: Toney, Jeffrey H.
  APPLICANT: Bruhn, Suzanne L.
  APPLICANT: Pil, Pieter M.
  APPLICANT: Brown, Steven
  APPLICANT: Kellett, Patti
  APPLICANT: Lippard, Stephen J.
  APPLICANT: Essigmann, John M.
  TITLE OF INVENTION: DNA Structure Specific Recognition
  TITLE OF INVENTION: Protein and Uses Therefor
  NUMBER OF SEQUENCES: 13
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
    STREET: 2 Militia Drive
    CITY: Lexington
    STATE: MA
    COUNTRY: USA
    ZIP: 02173
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: PCT/US92/11107
    FILING DATE: 19921218
    CLASSIFICATION:
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/539,906
    FILING DATE: 18-JUN-1990
  ATTORNEY/AGENT INFORMATION:
    NAME: Granahan, Patricia
    REGISTRATION NUMBER: 32,227
    REFERENCE/DOCKET NUMBER: MIT-4787AAA
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 617-861-6240
    TELEFAX: 617-861-9540
  INFORMATION FOR SEQ ID NO: 12:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 3292 base pairs
      TYPE: NUCLEIC ACID
      STRANDEDNESS: double
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    ORIGINAL SOURCE:
      ORGANISM: Saccharomyces cerevisiae
    IMMEDIATE SOURCE:
      CLONE: lambda yPt
    FEATURE:
      NAME/KEY: CDS
      LOCATION: 1..1626
```

; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1776 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36..1116
US-09-041-886-12

US-07-814-964-12

Query Match   3.68;   Score 42.8; DB 4; Length 1776;
Best Local Similarity 61.8%; Pred. No. 0.02; Mismatches 42; Indels 0; Gaps 0;
Matches 68; Conservative 0;

Qy   808 GCTGCTGCTGCGCGCCCCGTTCGTTGTTGTGGCGTCCAAGTGCTGCTGCCGTTGTTTCT 867
          |||||||||||||||||||||||||||||||| ||||||||| ||| | || |||||
Db   975 GCTGGCTGCTCGCTGCTGCCTGCGCTGGCTGCTGCTGCTGCCTGCGCTGGCTGCTTTT   916

Qy   868 GCTGTAGAAGAAAAAGAGGATTTCGTATTCAATTTCAAAAGAAATCTGAA 917
          |||| ||||||||||||||||||||| ||| ||||||||| |||||
Db   868 GCTGCTGTTTTTCAAAGTAGGCTTCTCGTCTTCCGAAGCTCTTCTGAA   915

RESULT 5
US-07-814-964-12/c
; Sequence 12, Application US/07814964
; Patent No. 5359047
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellett, Patti
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/814,964
; FILING DATE: 19911226
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3292 base pairs
; TYPE: NUCLEIC ACID

; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Saccharomyces cerevisiae
; IMMEDIATE SOURCE:
; CLONE: lambda yPt
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1626
US-07-814-964-12

Query Match   3.68;   Score 42.6; DB 1; Length 3292;
Best Local Similarity 64.9%; Pred. No. 0.031; Mismatches 34; Indels 0; Gaps 0;
Matches 63; Conservative 0;

Qy   804 ATACGCTGCGTGCCTGCGCCGCCGTTCGTTGTTGTGGCTGCTGCTGCCGTTGTTGT 863
          ||| ||||| | ||| ||| ||||||||||||||| ||||| || ||||| |||
Db   818 AGATGATGAGTGTTGTTGGTTGAAGTTGTTGTTGCAACTGGGTGTTGCTGTTGTAGC 759

Qy   864 TTCTGCTGTTAGAGAGAAAAAGAGGATTTCGTATTCAAT 900
          |||||||| ||||||||||||||||||||||| | ||
Db   758 TGCTGCTGTTGCTGCCAACTGCTGTTGTTGTTGTTGAATT 722

RESULT 6
US-08-258-442-12/c
; Sequence 12, Application US/08258442
; Patent No. 5670621
; GENERAL INFORMATION:
; APPLICANT: Donahue, Brian A.
; APPLICANT: Toney, Jeffrey H.
; APPLICANT: Bruhn, Suzanne L.
; APPLICANT: Pil, Pieter M.
; APPLICANT: Brown, Steven
; APPLICANT: Kellett, Patti
; APPLICANT: Essigmann, John M.
; APPLICANT: Lippard, Stephen J.
; TITLE OF INVENTION: DNA Structure Specific Recognition
; TITLE OF INVENTION: Protein and Uses Therefor
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/258,442
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/539,906
; FILING DATE: 18-JUN-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4787AAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3292 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

~ STREET: Two Militia Drive
~ CITY: Lexington
~ STATE: Massachusetts
~ COUNTRY: USA
~ ZIP: 02173-4799
~ COMPUTER READABLE FORM:
~ MEDIUM TYPE: Floppy disk
~ COMPUTER: IBM PC compatible
~ OPERATING SYSTEM: PC-DOS/MS-DOS
~ SOFTWARE: PatentIn Release #1.0, Version #1.30
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/08/531,927B
~ FILING DATE: 21-SEP-1995
~ CLASSIFICATION: 435
~ PRIOR APPLICATION DATA:
~ APPLICATION NUMBER: JP H6-251600
~ FILING DATE: 21-SEP-1994
~ ATTORNEY/AGENT INFORMATION:
~ NAME: Granahan, Patricia
~ REGISTRATION NUMBER: 32,227
~ REFERENCE/DOCKET NUMBER: ATH95-01A
~ TELECOMMUNICATION INFORMATION:
~ TELEPHONE: 617-861-6240
~ TELEFAX: 617-861-9540
~ INFORMATION FOR SEQ ID NO: 3:
~ SEQUENCE CHARACTERISTICS:
~ LENGTH: 325 base pairs
~ TYPE: nucleic acid
~ STRANDEDNESS: single
~ TOPOLOGY: linear
~ FEATURE:
~ NAME/KEY: modified_base
~ LOCATION: 145
~ FEATURE:
~ NAME/KEY: modified_base
~ LOCATION: 194
US-08-531-927B-3

Query Match
Best Local Similarity   57.8%;   Score 48.2;   DB 2;   Length 325;
Matches 100; Conservative 2; Mismatches 70; Indels 1; Gaps 1;
                4.1%;   Pred. No. 0.00023;

```
Qy  792 TGTACTCTTACAATACCGTCGCTGCCTGCCGCGTCGTTGTTGTGCCTGCAACTGCTGC 851
        |||  ||  || |||||||||||||||  |  |||||||  || |||||||||||||
Db  210 TGTCCTGATAGGTCCCCSCGTCGCTGCTGCTGCTGCCTGCTGCTGCTGCTGCTGCTGC 151

Qy  852 TGCCGTTGTTGTTGTTCTGCCTGTRAGAA-GAAAAAGAGGATTTCGTTGTTGCTGAGAA 910
        |||||||||||||||||||||| | ||||| | ||||| ||||||  ||  | ||||||
Db  150 TGCTGYTGCCTGCTCTTTTCGTGCYGTCTGAAACAATCAAAAGTGAAGTATATTAAAAACA 91

Qy  911 ATCTGAAAAAGAGAAGACAAACAAAAGAAAACTGAGACAGAAGTGGAAATGAAA 963
        ||||||  ||||  | |  | ||| || ||||||  |||||  || ||||||
Db   90 AAACTTAAAAGAATAAAATACACCATGAGAAAAAACTATTCATAAGGAAAAATACA 38
```

RESULT 3
US-08-531-927B-1/c
Sequence 1, Application US/08531927B
Patent No. 5840491
GENERAL INFORMATION:
~ APPLICANT: Kakizuka, Akira
~ TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
~ Patent No. 5840491
~ NUMBER OF SEQUENCES: 23
~ CORRESPONDENCE ADDRESS:
~ ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
~ STREET: Two Militia Drive
~ CITY: Lexington
~ STATE: Massachusetts
~ COUNTRY: USA
~ ZIP: 02173-4799
~ COMPUTER READABLE FORM:

~ MEDIUM TYPE: Floppy disk
~ COMPUTER: IBM PC compatible
~ OPERATING SYSTEM: PC-DOS/MS-DOS
~ SOFTWARE: PatentIn Release #1.0, Version #1.30
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/08/531,927B
~ FILING DATE: 21-SEP-1995
~ CLASSIFICATION: 435
~ PRIOR APPLICATION DATA:
~ APPLICATION NUMBER: JP H6-251600
~ FILING DATE: 21-SEP-1994
~ ATTORNEY/AGENT INFORMATION:
~ NAME: Granahan, Patricia
~ REGISTRATION NUMBER: 32,227
~ REFERENCE/DOCKET NUMBER: ATH95-01A
~ TELECOMMUNICATION INFORMATION:
~ TELEPHONE: 617-861-6240
~ TELEFAX: 617-861-9540
~ INFORMATION FOR SEQ ID NO: 1:
~ SEQUENCE CHARACTERISTICS:
~ LENGTH: 1776 base pairs
~ TYPE: nucleic acid
~ STRANDEDNESS: single
~ TOPOLOGY: linear
~ FEATURE:
~ NAME/KEY: CDS
~ LOCATION: 36..1115
US-08-531-927B-1

Query Match
Best Local Similarity   61.8%;   Score 42.8;   DB 2;   Length 1776;
Matches 68; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
                3.6%;   Pred. No. 0.02;

```
Qy  808 GCTGCTGCCTGCCGCGCCGTCGTTGTTGTGGCTGCAACTGCTGCTGCCGTTGTTGTTTCT 867
        |||||||||||| |||| |  ||| || || |  |||| |||||| ||||| ||| |||
Db  975 GCTGCTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGTTGCTGCTGCTTTT 916

Qy  868 GCTGTAGAAGAAAAAGAGGATTTCGTATTCAATTTCAAAAGAAAATCTGAA 917
        |||  ||||||   ||  |  |  |  || |  ||||  ||| ||||||
Db  915 GCTGCTGTTTTCAAAGTAGGCTTCTCGTCTGGAAGCTCTTCTGAA 866
```

RESULT 4
US-09-041-886-12/c
Sequence 12, Application US/09041886
Patent No. 6235872
GENERAL INFORMATION:
~ APPLICANT: Bredesen, Dale E.
~ APPLICANT: Rabizadeh, Shahroz
~ TITLE OF INVENTION: Proapoptotic Peptides, Dependence
~ TITLE OF INVENTION: Polypeptides and Methods of Use
~ NUMBER OF SEQUENCES: 72
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/09/041,886
~ FILING DATE:
~ CLASSIFICATION:
~ CORRESPONDENCE ADDRESS:
~ ADDRESSEE: Campbell & Flores LLP
~ STREET: 4370 La Jolla Village Drive, Suite 700
~ CITY: San Diego
~ STATE: California
~ COUNTRY: United States
~ ZIP: 92122
~ COMPUTER READABLE FORM:
~ MEDIUM TYPE: Floppy disk
~ COMPUTER: IBM PC compatible
~ OPERATING SYSTEM: PC-DOS/MS-DOS
~ SOFTWARE: PatentIn Release #1.0, Version #1.25
~ CURRENT APPLICATION DATA:
~ APPLICATION NUMBER: US/09/041,886
~ FILING DATE:
~ CLASSIFICATION:
~ ATTORNEY/AGENT INFORMATION:
~ NAME: Campbell, Cathryn A.
~ REGISTRATION NUMBER: 31,815
~ REFERENCE/DOCKET NUMBER: P-LJ 2626
~ TELECOMMUNICATION INFORMATION:

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.4_p5_4578

Run on: April 28, 2003, 23:25:03 ; Search time 49.1906 Seconds
            (without alignments)
            7325.486 Million cell updates/sec

Title: US-09-729-264-1
Perfect score: 1175
Sequence: 1 ctgtctgcccatctgaata..............gtaataacaactgtagtatag 1175

Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued_Patents_NA:*
               1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
               2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
               3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
               4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
               5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
               6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	Match DB	ID	Description
1	51.8	4.4	397	3	US-09-253-691-3	Sequence 3, Appli
2	48.2	4.1	325	2	US-08-531-927B-3	Sequence 3, Appli
3	42.8	3.6	1776	2	US-08-531-927B-1	Sequence 1, Appli
4	42.8	3.6	1776	4	US-09-041-886-12	Sequence 12, Appl
5	42.6	3.6	3292	4	US-07-814-964-12	Sequence 12, Appl
6	42.6	3.6	3292	1	US-08-258-442-12	Sequence 12, Appl
7	42.6	3.6	3292	1	US-08-328-809-7	Sequence 7, Appl
8	42.6	3.6	3292	5	PCT-US92-11107-12	Sequence 12, Appl
9	41.6	3.5	78	4	US-09-043-303-12	Sequence 12, Appl
10	41.6	3.5	203	4	US-09-043-303-7	Sequence 7, Appl
11	41.2	3.5	379	1	US-08-145-617-5	Sequence 5, Appl
12	41	3.5	3771	1	US-08-185-432-3	Sequence 3, Appl
13	41	3.5	3771	1	US-08-185-432-23	Sequence 23, Appl
14	41	3.5	5063	1	US-08-479-913E-1	Sequence 1, Appl
15	40.8	3.5	6177	4	US-09-091-042A-1	Sequence 1, Appl
16	40.8	3.5	7257	4	Patent No. 5496550	Sequence 1, Appli
17	40.4	3.4	484	6	US-09-817-180-1	Sequence 1, Appl
18	40.4	3.4	2674	6	US-08-457-273B-41	Sequence 41, Appl
19	40.4	3.4	10348	2	US-08-556-419-13	Sequence 13, Appl
20	40.4	3.4	10348	3	US-09-041-886-14	Sequence 14, Appl
21	40.4	3.4	10356	1	US-08-453-265-5	Sequence 5, Appl
22	40.4	3.4	10356	1	US-08-246-982A-5	Sequence 5, Appl
23	40.4	3.4	10366	1	US-08-453-265-5	Sequence 5, Appl
24	40	3.4	75	4	US-09-043-303-10	Sequence 10, Appl
25	40	3.4	78	4	US-09-043-303-11	Sequence 11, Appl
26	40	3.4	154	1	US-08-469-802B-6	Sequence 6, Appli
27	40	3.4	154	2	US-08-267-803B-6	Sequence 6, Appli
28	40	3.4	165	4	US-09-043-303-17	Sequence 17, Appl
29	40	3.4	168	1	US-08-469-802B-4	Sequence 4, Appli
30	40	3.4	168	2	US-08-267-803B-4	Sequence 4, Appli
31	40	3.4	171	2	US-08-469-802B-5	Sequence 5, Appli
32	40	3.4	171	2	US-08-267-803B-5	Sequence 5, Appli
33	40	3.4	195	1	US-08-469-802B-2	Sequence 2, Appli
34	40	3.4	195	1	US-08-267-803B-2	Sequence 2, Appli
35	40	3.4	234	1	US-08-469-802B-3	Sequence 3, Appli
36	40	3.4	234	1	US-08-267-803B-3	Sequence 3, Appli
37	40	3.4	270	4	US-09-146-054-8	Sequence 8, Appli
38	40	3.4	270	4	US-09-664-977A-8	Sequence 8, Appli
39	40	3.4	477	4	US-09-135-994-1	Sequence 1, Appli
40	40	3.4	506	2	US-08-469-802B-7	Sequence 7, Appli
41	40	3.4	506	2	US-08-267-803B-7	Sequence 7, Appli
42	40	3.4	623	4	US-08-469-802B-2	Sequence 2, Appli
43	40	3.4	3366	1	US-08-469-802B-1	Sequence 1, Appli
44	40	3.4	3366	2	US-08-267-803B-1	Sequence 1, Appli
45	40	3.4	3376	1	US-08-320-559-29	Sequence 29, Appl

ALIGNMENTS

RESULT 1
US-09-253-691-3/c
; Sequence 3, Application US/09253691
; Patent No. 6124100
; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; TITLE OF INVENTION: Using Trinucleotide Repeats Sequence
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; CURRENT FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
US-09-253-691-3

Query Match   4.4%; Score 51.8; DB 3; Length 397;
Best Local Similarity   58.9%; Pred. No. 2.4e-05;
Matches   89; Conservative   0; Mismatches 62; Indels   0; Gaps   0;

Qy   808 GCTGCTGCTGCTGCCCGGCGTCGTTGTGTGGCTGCAACTGCTGCCCGGTTGTTGTTTCT 867
              ||||||||||||||| |||||| |||||| |||||||||||||||||| ||||| ||
Db   200 GCTGCTGCTGCTGCTGCTGCTGCTG------GCTGCAACTGCTGCTGCTGCTGCTGCTGCT 141

Qy   868 GCTGTAGAAGAAAAAAGAGGAGATTTCGTATTCAAATTTCAAAAGAAATCTGAAAAAGAGAAGA 927
              ||||| |||| |||||| || ||||| ||| |||||| ||| |||||||||||||||||||
Db   140 GCTGTTGCTGCTGCTTTTGCTGCTGCTGCTGCTGAAACATTCAAAAGTGAAGTATATTTAAAAAACAA 81

Qy   928 CAAACAAAGAAAACTGAGACAGAAAAGTGGAAA 958
              ||| || |||||||| ||| ||| |||||||
Db   80 AACTTAAAAGAATAAAATACACCATGAGAAAA 50

RESULT 2
US-08-531-927B-3/c
; Sequence 3, Application US/0853192 7B
; Patent No. 5840491
; GENERAL INFORMATION:
; APPLICANT: Kakizuka, Akira
; TITLE OF INVENTION: DNA Sequence Encoding the Machado-Joseph
; Patent No. 5840491
; TITLE OF INVENTION: Disease Gene and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

```
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 29
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 33
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 21
US-09-864-761-2182

Query Match         3.7%;   Score 43.2;   DB 10;   Length 459;
Best Local Similarity   79.7%;   Pred. No. 0.0067;
Matches  51;   Conservative  0;   Mismatches  13;   Indels  0;   Gaps  0;

Qy    808 GCTGCTGCTGCTGCCGCCCGTCGTTGTTGTTGGGCTGCCAAACTGCTTGCTGCCCGTTGTTTGTTTTCT 867
          ||||||||||||||||| || | | | | ||| | ||||||||| | ||||||||||||||||||
Db    389 GCTGCTGCTGCTGCTGTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGTTGTTGTTGGCT 330

Qy    868 GCTG 871
          ||||
Db    329 GCTG 326

RESULT 15
US-09-416-384A-6/c
Sequence 6, Application US/09416384A
Patent No. US20020081584A1
;   GENERAL INFORMATION:
;   APPLICANT:   BLUMENFELD, Marta
;   APPLICANT:   BOUGUELERET, Lydie
;   APPLICANT:   CHUMAKOV, Ilya
;   APPLICANT:   COHEN, Daniel
;   APPLICANT:   ESSIOUX, Laurent
;   TITLE OF INVENTION: Genes, proteins and biallelic markers related to central...
;   FILE REFERENCE: GENSET.045AUS
;   CURRENT FILING DATE:   1999-10-12
;   CURRENT APPLICATION NUMBER: US/09/416,384A
;   PRIOR APPLICATION NUMBER: 60/106,457
;   PRIOR FILING DATE:   1999-10-30
;   PRIOR APPLICATION NUMBER: 60/103,955
;   PRIOR FILING DATE:   1998-10-12
;   PRIOR APPLICATION NUMBER: 60/132,277
;   PRIOR FILING DATE:   1999-05-03
;   NUMBER OF SEQ ID NOS:   71
;   SEQ ID NO 6
;   LENGTH: 1791
;   TYPE: DNA
;   ORGANISM: mus musculus
US-09-416-384A-6

Query Match         3.7%;   Score 43.2;   DB 10;   Length 1791;
Best Local Similarity   75.0%;   Pred. No. 0.017;
Matches  54;   Conservative  0;   Mismatches  18;   Indels  0;   Gaps  0;

Qy    807 CGCTGCTGCTGCTGCCCGCGCGTCGTTGTTGTTGGCGTGCGCAACTGCTGCCGTTGTTGTTTTC 866
          ||||||||||||||||||| |||||| ||||||| | ||| || |||||| | ||||||||| |||
Db    336 CGCTGCTGCTGCTGCTGCTGCTGCTGTTGCTGCTGTTGCTGCTGCTGCTGCTGTTGCTGCTGC 277

Qy    867 TGCTGTAGAAGA 878
          ||||||| | ||
Db    276 TGCTGTTGGTGA 265

Search completed: April 28, 2003, 20:36:16
Job time : 2140.37 secs
```

APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 18923
LENGTH: 293
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC009954.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 47
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 53
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 69
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 27
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 16
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 21
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 45
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 29
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 33
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 21
OTHER INFORMATION: SWISSPROT HIT: P53360, EVALUE 4.00e-15
OTHER INFORMATION: NT HIT: Z22828.1, EVALUE 9.00e-59
OTHER INFORMATION: EST_HUMAN HIT: AW409956.1, EVALUE 1.00e-58

US-09-864-761-18923

Query Match   79.7%; Score 43.2; DB 10; Length 293;
Best Local Similarity   79.7%; Pred. No. 0.005;
Matches 51; Conservative   0; Mismatches 13; Indels   0; Gaps   0;

Qy   808 GCTGCTCGTCGTCGCCGCCGCCGTGCTTGTTGTGGCTGCAACTGCCTGTTGTTGTTTCT 867

RESULT 14
US-09-864-761-2182/c
Sequence 2182, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

SEQ ID NO 2182
LENGTH: 459
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC009954.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 47
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 53
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 69
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 27
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 16
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 21
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 45

```
Qy    808 GCTGCTGCTGCTGCCGCGTCGTTGTTGTGGCTGCAAGTGCTGCTTGCCGTTGTTGTTTCT 867
          |||| |||||||| |||||||||||||||||||||  ||||||||  ||||||| ||
Db   6175 GCTGTTGCTGCTGCTCGTTGCTGGCTGCTGCTGCTCGTTGCTGCTGCTGTTGCTGTTGCT 6116

Qy    868 GCTGTAGAAGAA 879
          |||| |||
Db   6115 GCTGCTGCCGCA 6104
```

RESULT 10
US-09-728-444-151
; Sequence 151, Application US/09728444
; Patent No. US20020161207A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020161207A1e1 Murine Polynucleotide Sequences
; TITLE OF INVENTION: and Mutant Cells and Mutant Animals Defined Thereby
; FILE REFERENCE: LEX-0100-USA
; CURRENT APPLICATION NUMBER: US 09/728,444
; CURRENT FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: US 60/168,360
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 1206
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 151
; LENGTH: 455
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(455)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-444-151

Query Match   3.7%; Score 44; DB 9; Length 455;
Best Local Similarity   69.4%; Pred. No. 0.0038;
Matches   59; Conservative   0; Mismatches   26; Indels   0; Gaps   0;

```
Qy    810 TGCTGCTGCTGCCGCGTCGTTGTTGTGGCTGCAACTGCTGCTCGTCCGTTGTTGTTTCTGC 869
          |||||||||||| |||||||||||||  |||||||  ||||  ||||  ||  |||   ||
Db    162 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCGTGCTCGTGCTCGTGCTGCTGCTGC 221

Qy    870 TGTAGGAAGAAAAAGAGGAGGAATTTCGTA 894
          ||||  ||| |||
Db    222 TGCAAACCTANACCAGGATGTGGGA 246
```

RESULT 11
US-09-738-626-544
; Sequence 544, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 544
; LENGTH: 384
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-544

Query Match   3.7%; Score 43.4; DB 9; Length 384;
Best Local Similarity   64.4%; Pred. No. 0.0052;
Matches   65; Conservative   0; Mismatches   36; Indels   0; Gaps   0;

```
Qy    806 ACGGTCGTGCTGTGCTGCCGCGCCGTCGTTGTTGTGGCTGCAACTGCTGCTCGTTGTTGTTT 865
          |||  ||| |||| ||||||||  |||||  |||||||  |  |||||||  ||||||||
Db     98 ACGTAACCGCAGACTCGTCCAGTTGCAGTTGCAGTTGCTGCTGCAGGCGCGCAGGCGCGAAGCTG 157

Qy    866 CTGCTGTAGAAGAAAAAGAGGAGGATTTCGTATTCAAATTCAAA 906
          |||||| |||| | | |||||  ||  ||||||  |  ||||||
Db    158 CTGCTGCAGAGAGAGAAGAGAAGGACGAGTTCGACGTCGTTCTCGAA 198
```

RESULT 12
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match   3.7%; Score 43.4; DB 9; Length 3309400;
Best Local Similarity   65%; Pred. No. 2.1;
Matches   65; Conservative   0; Mismatches   36; Indels   0; Gaps   0;

```
Qy    806 ACGGTCGTGCTGTGCTGCCGCGTCGTTGTTGTGGCTGCAACTGCTGCTCGTTGTTGTTT 865
          |||  ||| ||||| |||||||  |||||  ||||||||  |  |||||||  ||||||||
Db 510688 ACGTAACGCAGCTGCTCCAGTTGCAGTTGCAGTTGCTGCTGCAGGCGCGTTGTTGTTT 510747

Qy    866 CTGCTGTAGAAGAAAAAGAGGAGGATTTCGTATTCAAATTCAAA 906
          |||||  |||| | | |||||  ||  ||||||  |  ||||||
Db 510748 CTGCTCCAGAAGAGAAGAGAAGAGGACGAGTTCGACGTCGTTCTCGAA 510788
```

RESULT 13
US-09-864-761-18923/c
; Sequence 18923, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:

DR   WPI; 2000-662424/64.
XX
PT   Plasmid vector arraying neuropsychiatric gene containing trinucleotide repeats its use more than copy number
PT   containing trinucleotide repeats its use thereof -
XX
PS   Disclosure; Page 12; 23pp; Korean.

CC   This invention relates to the use of a plasmid vector array
CC   containing a neuropsychiatric gene containing trinucleotide repeats.
CC   Tri-nucleotide repeats and their expansion areknown to be involved
CC   in a number of diseases including fragile X syndrome, spinocerebellar
CC   ataxia type III, hereditary hypertrophic, cardiomyopathy, dentatorubral
CC   and pallidoluysian atrophy (DRPLA), spinocerebellar ataxia, Marfan
CC   syndrome; X-linked spinobubar atrophy. The present sequence
CC   represents a nucleotide sequence of the invention showing the
CC   the CAG tri-nucleotide repeats.
XX
SQ   Sequence 403 BP; 121 A; 108 C; 105 G; 69 T; 0 other;

Query Match   4.2%; Score 51.8; DB 21; Length 403;
Best Local Similarity   58.9%; Pred. No. 0.00042;
Matches   89; Conservative   0; Mismatches   62; Indels   0; Gaps   0;

Qy   873 GCTGCTGCTGCTGCTGCGCCGTCGTTGTGTTGTGGCTGCCAACTGCTGCCTTGTTGTTTCT 932
         |||||||||||||||||||||| |||||||||| ||||||||| ||||||||||||||||
Db   215 GCTGCTGCTGCTGCTGCTGCTGCTGCCTGCTGCTGCGCTGCTGCCTGCTGTCTGCTGCGCT 156

Qy   933 GCTGTAGAAGAAAAGAGAGAGAGATTTCGTATTCAATTTCAAAAGAAATCTGAAAAGAGAGAGA 992
         ||||| ||||||| ||| ||| ||||| || ||||| ||||||||||||||| || |||||
Db   155 GCTGTTGCTGCTCCTTTTGCTGCTGCTGTGCTGAAACATTCAAAAAGTGAAGTATATTAAAAACAA 96

Qy   993 CAAACAAAGAAAACTGAGACACAGAAAAGTGGAAA 1023
         ||| ||| ||| || || |||| |||
Db   95 AACTTAAAAGAATAAAATACACCATGAGAAAA 65

CC   human neuroblastoma. The nucleic acids are applicable as a probe or
CC   primer in diagnosing the prognosis of human neuroblastoma, malignancy and
CC   susceptibility indicators or tumour markers for anti-cancer agents. The
CC   gene information for diagnosing prognosis is related to factors similar
CC   to that for N-myc and TrkA genes.
XX
SQ   Sequence 742 BP; 180 A; 183 C; 152 G; 219 T; 8 other;

Query Match   4.0%; Score 49.6; DB 22; Length 742;
Best Local Similarity   69.8%; Pred. No. 0.0024;
Matches   67; Conservative   0; Mismatches   29; Indels   0; Gaps   0;

Qy   23 GTGGATCAGTTTCCTAGGCTGCCATAACAAAAGCACCATAACCTGGTGGCTTAGAACAATG 82
         || | ||||| ||||| |||||||| ||| ||||||||||||| | ||||| ||||||
Db   343 GTATATTAGTTTATAAGGCTCCCATAACAAAAGCACCACAGACTGGGTGGCTTAAACAAAA 402

Qy   83 GAAAGGCATTGCTCACGGTTCCAGAAGCTTGTAGGT 118
         |||| |||||||| |||| |||| || |||| |||| ||
Db   403 GAAATCTATTTTCTCACACAGTTCTGGAGGCTAGAAGT 438

Search completed: April 29, 2003, 11:05:46
Job time : 322.05 secs

RESULT 15
AAI95221
ID   AAI95221 standard; cDNA; 742 BP.
XX
AC   AAI95221;
XX
DT   13-NOV-2001 (first entry)
XX
DE   Human neuroblastoma expressed polynucleotide SEQ ID NO 1296.
XX
KW   Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.
XX
OS   Homo sapiens.
XX
PN   WO20016G119-A1.
XX
PD   13-SEP-2001.
XX
PF   02-MAR-2001; 2001WO-JP01629.
XX
PR   07-MAR-2000; 2000JP-0159195.
XX
PA   (CHIB-) CHIBA PREFECTURE.
PA   (HISM) HISAMITSU PHARM CO LTD.
XX
PI   Nakagawara A;
XX
DR   WPI; 2001-565584/63.
XX
PT   Nucleic acids originating in gene expressed in human neuroblastoma,
PT   useful as probe or primer in diagnosing prognosis of human
PT   neuroblastoma, malignancy and susceptibility indicator or tumour marker
PT   for anti-cancer agents -
XX
PS   Claim 1; Page 977-978; 2979pp; Japanese.

CC   The invention relates to novel genes (AAI93926-AAI97963) expressed in

XX
CC This invention relates to the use of a plasmid vector array
CC containing a neuropsychiatric gene containing trinucleotide repeats.
CC Tri-nucleotide repeats and their expansion are known to be involved
CC in a number of diseases including fragile X syndrome, spinocerebellar
CC ataxia type III, hereditary hypertrophic, cardiomyopathy, spinocerebellar
CC and pallidoluysian atrophy (DRPLA), spinocerebellar ataxia, Marfan
CC syndrome; X-linked spinobubar atrophy. The present sequence represents
CC of the invention. The present sequence #5 disclosed in the scope
CC a tri-nucleotide repeat related sequence #5 disclosed in the scope
CC Note; Nucleotides which were illegible in the specification have
CC been represented by an n to keep nucleotide numbering correct.
XX
SQ Sequence 389 BP; 107 A; 92 C; 91 G; 93 T; 6 other;

Query Match   4.2%; Score 51.8; DB 21; Length 389;
Best Local Similarity 57.4%; Pred. No. 0.00042;
Matches   89; Conservative   0; Mismatches 66; Indels   0; Gaps   0;

Qy 873 GCTGCTGCTGCTGCCGCCCGTCGTTGTTGTGGCTGCAACTGCTGCCGTTGTTGTTTCT 932
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 178 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCCTGCTGCTGCTGCTGCTGCTT 119

Qy 933 GCTGTAGAGAGAAAAAAGAGGATTTCGTATTCAAATTTCAAAAAAGAAATCTGAAAAAAGAGAGA 992
         ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 118 GCTGTTGCTGCTGCTTTTTGCTCGTGTCTGAAACATTCAAAAAGTGAAGTATATATTTAAAAACA 59

Qy 993 CAAACAAAGAAAACTGAGACAGAAAAGTGGAAAATGAA 1027
         ||| ||| ||| ||||| || | | ||
Db   58 AANTTAAAAAGAANAACCTACACCATGAGNTANAAAAA 24

RESULT 13
ID   AAX89891/c
ID   AAX89891/c
AC   AAX89891;
XX
DT   05-NOV-1999 (first entry)
XX
DE   Spinocerebellar ataxia type III (SCAIII) gene fragment.
KW   Spinocerebellar ataxia type III; SCAIII; reverse dot hybridisation;
KW   PCR-microplate hybridisation; PCR-MPH; trinucleotide repeat; TNR;
KW   SCAIII syndrome; ss.
XX
OS   Homo sapiens.
XX
FH
FT   Key   Location/Qualifiers
FT   repeat_region   137..355
FT   /*tag= a
FT   repeat_unit   137..139
FT   /*tag= b
FT   /note= "trinucleotide repeat"
XX
XX
PN   WO9943852-A1.
XX
PD   02-SEP-1999.
XX
PF   18-FEB-1999; 99WO-KR00078.
XX
PR   26-FEB-1998; 98KR-0006278.
XX
PA   (JIND/) JIN D K.
PA   (SMSU) SAMSUNG FINE CHEM CO LTD.
XX
PI   Jin DK;
XX
DR   WPI; 1999-527634/44.
PT   Diagnosis of spinocerebellar ataxia type III (SCA III) syndrome
PT   using techniques which ensure highly accurate diagnosis

PS   Claim 1; Page 12-13; 28pp; English.
XX
CC   The invention relates to the diagnosis of spinocerebellar ataxia type
CC   III (SCAIII) syndrome using reverse dot hybridisation or PCR-microplate
CC   hybridisation (PCR-MPH). The method comprises attaching a potion of the
CC   SCAIII gene containing 73 copies of the trinucleotide (CAG) repeat unit
CC   (the present sequence) to a substrate, and hybridising with amplified
CC   tested genomic DNA containing copies of the trinucleotide units, PCR
CC   amplified with labeled primers (AAX89889-90). The new method is useful
CC   for diagnosis of SCAIII syndrome, and for determining the severity of
CC   the disease. The present sequence represents the SCAIII gene fragment
CC   containing 73 trinucleotide (TNR) repeats.
XX
SQ   Sequence 397 BP; 124 A; 104 C; 65 G; 104 T; 0 other;

Query Match   4.2%; Score 51.8; DB 20; Length 397;
Best Local Similarity 58.9%; Pred. No. 0.00042;
Matches   89; Conservative   0; Mismatches 62; Indels   0; Gaps   0;

Qy 873 GCTGCTGCTGCTGCCGCCCGTCGTTGTTGTGGCTGCAACTGCTGCCGTTGTTGTTTCT 932
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 200 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCCTGCTGCTGCTGCTGCTGCTT 141

Qy 933 GCTGTAGAGAGAAAAAAGAGGATTTCGTATTCAAATTTCAAAAAAGAAATCTGAAAAAAGAGAGA 992
         ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 140 GCTGTTGCTGCTGCTTTTGCTCGTGTCTGAAACATTCAAAAAGTGAAGTATATTTAAAAAAACA 81

Qy 993 CAAACAAAGAAAACTGAGACAGAAAAGTGGAAA 1023
         ||| ||| |||| | ||| || | | |||
Db   80 AACTTAAAAGAATAAATACCACCATGAGAAAA 50

RESULT 14
ID   ABKI0240/c
ID   ABKI0240/c
AC   ABKI0240;
XX
DT   20-MAY-2002 (first entry)
XX
DE   Trinucleotide repeat sequence #3.
KW   Trinucleotide repeat; fragile X syndrome; ds;
KW   spinocerebellar ataxia type III; DRPLA; neuropsychiatric;
KW   hereditary hypertrophic cardiomyopathy; Marfan syndrome;
KW   dentatorubral and pallidoluysian atrophy;
KW   spinocerebellar ataxia; X-linked spinobubar atrophy.
XX
OS   Homo sapiens.
XX
FH
FT   Key   Location/Qualifiers
FT   repeat_region   152..347
FT   /*tag= a
FT   repeat_unit   152..154
FT   /rpt_type= TANDEM
FT   /*tag= b
FT   /note= "CAG type repeat"
XX
XX
PN   KR200000003004-A.
XX
PD   15-JAN-2000.
XX
PF   25-JUN-1998; 98KR-0024064.
XX
PR   25-JUN-1998; 98KR-0024064.
XX
PA   (SMSU) SAMSUNG MEDICAL CENT.
PA   (JIND/) JIN D G.
XX
PI   Jin DG;
XX

CC   and pallidoluysian atrophy (DRPLA), spinocerebellar ataxia, Marfan
CC   syndrome; X-linked spinobubar atrophy. The present sequence
CC   represents a nucleotide sequence of the invention showing the
CC   the CAG tri-nucleotide repeats.
XX
SQ   Sequence 277 BP; 80 A; 66 C; 62 G; 69 T; 0 Other;

Qy   Query Match   4.3%;   Score 53.4;   DB 21;   Length 277;
  Best Local Similarity   59.6%;   Conservative   0;   Mismatches   61;   Indels   0;   Gaps   0;
  Matches   90;   Conservative   0;   Pred. No. 0.00013;

Qy   873 GCTGCTGCTGCTGCTGCCGCGCGCGCCCGTCGTTGTGTTGGCTGCAACTGCGTTGTTGTTTCT 932
        ||||||||||||||||| |||||||||| ||||||||||| || ||||||||||||| |||
Db   215 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 156

Qy   933 GCTGTAGAGAAAAAAGAGGATTTCGTATTCAATTTCAAAAAGAGAGAAGA 992
        ||||| |||||||||| ||||||| ||| ||||| |||||||||
Db   155 GCTGTTGCTGCTTTTGCTGCTGTCTGAAACATTCAAAAGTGAAGTATATTTTAAAAAACAA 96

Qy   993 CAAACAAAGAAAACTGAGACAGAAAAGTGGAAA 1023
        ||| ||| | ||| ||||| |||
Db   95 AACTTAAAAGAATAAAATACACCATGAGAAAA 65

RESULT 11
ID   ABL29756 standard; DNA; 7029 BP.
XX
AC   ABL29756;
XX
DT   26-MAR-2002 (first entry)
XX
PN   WO200171042-A2.
XX
OS   Drosophila melanogaster.
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 40741.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New lsolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 40741; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7029 BP; 1426 A; 1440 C; 1945 G; 2218 T; 0 other;

Qy   Query Match   4.3%;   Score 52.8;   DB 23;   Length 7029;
  Best Local Similarity   69.2%;   Pred. No. 0.001;
  Matches   72;   Conservative   0;   Mismatches   32;   Indels   0;   Gaps   0;

Qy   873 GCTGCTGCTGCTGCTGCCGCGCGCGCCCGTCGTTGTGTTGGCTGCAACTGCGTTGTTGTTTCT 932
        ||||| ||| ||||||| |||||||| || |||||||||| |||| ||||||||| ||||
Db   5500 GCTGTAGAGAAAAAAGAGGCTGCTGCTGCTGCGCTCGCTGGCTGCTTGTTGTTGTTGTTGCT 5559

Qy   933 GCTGTAGAGAAAAAAGAGGATTTCGTATTCAATTTCAAAAAGAGAGAA 976
        ||||| ||| |||||||| ||||| || ||| ||| || ||| | |||
Db   5560 GCTGTATTTGCTAAAAAGAAGTAATCATTAATTTCCGGTAAAAA 5603

RESULT 12
ID   ABK10248 standard; DNA; 389 BP.
XX
AC   ABK10248;
XX
DT   20-MAY-2002 (first entry)
XX
DE   Tri-nucleotide repeat related sequence #5.
XX
KW   Trinucleotide repeat; fragile X syndrome; ds;
KW   spinocerebellar ataxia type III; DRPLA; neuropsychiatric;
KW   hereditary hypertrophic cardiomyopathy; Marfan syndrome;
KW   dentatorubral and pallidoluysian atrophy;
KW   spinocerebellar ataxia; X-linked spinobubar atrophy.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   misc_difference   28
FT   /*tag= a
FT   /note= "Illegible in the specification"
FT   misc_difference   31
FT   /*tag= b
FT   /note= "Illegible in the specification"
FT   misc_difference   46
FT   /*tag= c
FT   /note= "Illegible in the specification"
FT   misc_difference   56
FT   /*tag= d
FT   /note= "Illegible in the specification"
FT   misc_difference   203
FT   /*tag= e
FT   /note= "Illegible in the specification"
FT   misc_difference   276
FT   /*tag= f
FT   /note= "Illegible in the specification"
XX
OS   Unidentified.
XX
PN   KR200003004-A.
XX
PD   15-JAN-2000.
XX
PF   25-JUN-1998; 98KR-0024064.
XX
PR   25-JUN-1998; 98KR-0024064.
XX
PA   (SMSU) SAMSUNG MEDICAL CENT.
PA   (JIND) JIN D G.
XX
PI   Jin DG;
XX
DR   WPI; 2000-662424/64.
XX
PT   Plasmid vector arraying neuropsychiatric gene more than copy number
PT   containing trinucleotide repeats its use thereof -
XX
PS   Disclosure; Fig 10; 23pp; Korean.

```
XX
PF   28-JUN-2001;
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75545.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 1; Fig 6; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of B7-L
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of mouse B7_L_m2.
XX
SQ   Sequence 895 BP; 250 A; 223 C; 212 G; 210 T; 0 other;

Query Match             15.0%;  Score 185.6;  DB 24;  Length 895;
Best Local Similarity   73.8%;  Pred. No. 5.8e-41;
Matches  236;  Conservative   0;  Mismatches  84;  Indels   0;  Gaps   0;

Qy   115 AGGTTCTGGGTCTTGGTAATGAAGTCATAGAAGGCCCCCAGAANTGCAACAGTCCTGRAAGG 174
         ||||||| | |||| | |||||||||||||||| | ||||||||||| |||| | || |||| ||
Db   109 AGCTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAANTGTAACAGTCCTAAAGGA 168

Qy   175 CTCCCAGGCCTCCGTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCAIGTGGGGTCT 234
         || |||||||||| ||| |||||||||| | ||||||| |||||||||| ||| || |||||
Db   169 CTCAGAGGCTCACTTCAACTGCACCGTCTCCGTGAAGCTTCTCAITGTGGGACTCT 228

Qy   235 CACTGACATGGTGGTTGGTCAAGGCGTCAGGCCCATGGACCCATCATACCAATGACCGCTT 294
         ||||| ||||| |||||| || |||||||||||| | |||| |||||||||| ||||||||
Db   229 TAACCAAATGGGTGGTGCTGAGTCTCACCACCCAAGGACCCAATCAITCAICAACAACCGCTT 288

Qy   295 CACCTCTCAGAGGTACGACAGGGGGGGAACTTCACCTCGGACAGATGATCATCCACAATGT 354
         |||| | ||||||||| |||| |||||| ||||||||| || ||| |||| |||| |||||
Db   289 CACCTATGCCAGTTACACAGCACTGACAGCTTCAICTCGGAGTTGAICAICCAITGAIGT 348

Qy   355 GGAGCCCAGTGAIITCGGGGAACATCAIGAIGCAGCCTCAIGAIACAIGTCGGCCTGCAIGGAIC 414
         || |||||||||| || ||||||||||| | ||||||||||| |||||||| | || || |||
Db   349 GCAGCCCAITGAICTCGGGGAITCCGTGCCAITGCCAITGCGCAIGAIACAITCCAIGGTTTGGAITC 408

Qy   415 TGCTTACCTTACCGTCCAAG 434
         ||||||||||||||||||||
Db   409 TGCCTTCCTCTCCAITGTGCCAIAG 428
```

```
RESULT 10

ID   ABK10239/c
XX
AC   ABK10239;
XX
DT   20-MAY-2002 (first entry)
XX
DE   Trinucleotide repeat sequence #2.
XX
KW   Trinucleotide repeat; fragile X syndrome; ds; DRPLA;
KW   spinocerebellar ataxia type III; Marfan syndrome;
KW   hereditary hypertrophic cardiomyopathy; neuropsychiatric;
KW   dentatorubral and pallidoluysian atrophy;
KW   spinocerebellar ataxia; X-linked spinobubar atrophy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   repeat_region   152..194
FT                   /*tag= a
FT                   /rpt_type= TANDEM
FT   repeat_unit     152..154
FT                   /*tag= b
FT   repeat_region   198..222
FT                   /note= "CAG type repeat"
FT                   /*tag= a
FT   repeat_unit     198..200
FT                   /rpt_type= TANDEM
FT                   /*tag= b
FT                   /note= "CAG type repeat"
XX
CC
Homo sapiens.

KR2000003004-A.

15-JAN-2000.

25-JUN-1998;  98KR-0024064.

25-JUN-1998;  98KR-0024064.

(SMSU )  SAMSUNG MEDICAL CENT.
(JIND/)  JIN D G.

WPI; 2000-662424/64.

Jin DG;

Disclosure; Page 12; 23pp; Korean.

Plasmid vector arraying neuropsychiatric gene more than copy number
containing trinucleotide repeats its use thereof -

This invention relates to the use of a plasmid vector array
containing a neuropsychiatric gene containing trinucleotide repeats.
Tri-nucleotide repeats and their expansion areknown to be involved
in a number of diseases including fragile X syndrome, spinocerebellar
ataxia type III, hereditary hypertrophic, cardiomyopathy, dentatorubral
```

RESULT 8

```
ABK13034
ID   ABK13034 standard; cDNA; 754 BP.
XX
AC   ABK13034;
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding mouse B7-like protein, B7-L_m3.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic;
KW   antiinflammatory; dermatological; antipsoriatic; antirheumatic;
KW   antidiabetic; nephrotropic; antibacterial; virucide; neuroprotective;
KW   antiasthmatic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Mus musculus.
XX
FH   Key            Location/Qualifiers
FT   CDS            53..724
FT                  /*tag= a
FT                  /product= "B7-like protein, B7_Lm3"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75546.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 1; Fig 7; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusion, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of mouse B7-L_m3.
XX
SQ   Sequence 754 BP; 220 A; 191 C; 175 G; 168 T; 0 other;
```

```
Query Match             15.0%;  Score 185.6;  DB 24;  Length 754;
Best Local Similarity   73.8%;  Pred. No. 5.3e-41;
Matches  236;  Conservative   0;  Mismatches  84;  Indels   0;  Gaps  0;

Qy    115 AGGTTCTTCGGTCTGGTAATGAAGTCATAGAAGAGGACCCCAGAAGAATGCAACAGTCGTT 174
          || |||| || | || |||| | ||||||||||||   |||| || ||  ||||||||
Db    109 AGCTTCGGGATCCAGTTATCAGATCATAGAAGAGGTCTTCACAGAAGTAACAGTCCTAAGGA 168

Qy    175 CTCCCAGGCTGCCTTCGGTTCAACTGCACCGGTCTCCCAGGGCTGGGAAGGCTCATCATGT 234
          |||| || || |||||   ||||||||||||||  | || ||||  ||||| ||||||
Db    229 TAACCAAAATGGTGGGTGTCTGAGTCTCACCACCCAAGGGACCCATCATCACCACAACCGCTT 288

Qy    235 CAGTGACATGGTGGTGGTGCTAAGCGTGTCAGGCCCCATGGACCCATCATCACCAATGACCGCTT 294
          ||| ||||| | | || |||| ||  || ||||||||| ||||| ||||||||||||||||||||
Db    289 CACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATCCATGATGT 348

Qy    295 CACCTCTCAGAGGTGACGACCAGGCGGGGAACTTCACCTCGGGAGATGATCATCCACAATGT 354
          |||| |||||||| || || || || |||||||||| |||  ||||||||||| | |||||
Db    169 CTCAGAGCGCTCACTTCAACTGCACCGGTGGAAGCTTCATCATGTTGGACTTCT 228

Qy    355 GGAGCCCAGTGATTCGGGGGAAACATCAGATGCAGCCTCAGACACAGTCGCCTCGCATGGATC 414
          || || ||  || | ||||||||||||||| || || || ||||| |||| ||| ||||| |
Db    349 GCAGCCCAGTGACTCGGGGGATCCGTGCAATGCAGCGCTGCAGAGAACAGCCATGGGTTTGGATC 408

Qy    415 TGCTTACCTTACCGTCCAAG 434
          ||||| || |||||| ||||
Db    409 TGCCTTCCTCTCAGTGCAAG 428
```

RESULT 9

```
ABK13033
ID   ABK13033 standard; cDNA; 895 BP.
XX
AC   ABK13033;
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding mouse B7-like protein, B7-L_m2.
XX
KW   Mouse; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic;
KW   antiinflammatory; dermatological; antipsoriatic; antirheumatic;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Mus musculus.
XX
FH   Key            Location/Qualifiers
FT   CDS            53..865
FT                  /*tag= a
FT                  /product= "B7-like protein, B7-L_m2"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
```

Qy 115 AGGTTCTGGGTCTGGTAATGAAGTCAATAGAAGGCCCCAGAAATGCAACAGTCCTGAAGGG 174
       |||||| | |||||||||||||||||||||||||||| ||| || ||||||||||| ||||
Db 109 AGCTTCCGGATCCAGTTATCAGATCAATAGAAGGTCCTCAGAGAATGTAACAGTCCTAAAGGA 168

Qy 175 CTCCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGGAAGCTTCATCATGTGGGCTCT 234
Db 169 CTCAGAGGCTCACTTCAACTGCACCGTGACTCACGGTTCTCAGTAGTTCATGTGGACTCT 228

Qy 235 CAGTGACATGGTGGTGCTTAGGCGTCAGGCCCATGGAGGCGCCCATCATCACCAATGACCGCTT 294
Db 229 TAACCAAATGGTTGGTGCTGAGTCTCACCACCCAAGGACCCATCATCACCAACAACCGCTT 288

Qy 295 CACCTCTCAGAGGTACGACCAGGCGGGGGAACTTCACCTCTGGAGATGACATGATCATCCACAATGT 354
Db 289 CACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGAATCATCCATGATGT 348

Qy 355 GGAGCCCAGTGATTCGGGGAACATCAGATTGCACGCCTCCAGAACAGTGCGTTCTGCCATGGATC 414
Db 349 GCAGCCCAGTGACTCGGGATCCGGTGCAATGCAGCCTGCAGAACAGCCATGGGTTTGGATC 408

Qy 415 TGCTTACCTTACCGTCCAAGTTATGGAGGAGCTGTTCATTCCCAGTGTTAAATCTTGTAGT 474
Db 409 TGCCTTCCTCTCAGTGCAAGTCATGGGGGACCCTGAAACATTCCTAGCAACAAACCTTATAGT 468

Qy 475 CGCTGAGAATGAAACCTTGTGAAGTTACTTTGTCTTACCCTTCACACTGGACCGGCTCCCGGA 534
Db 469 CACTGAGGGTGAAACCTGTAATGTGACTTGCTATGCCGTGGGGCTGGAACTTCACTCCCGGA 528

Qy 535 TATTTCCTGGGAGCTCGGTCGTCCTCCTGGGTCCAGCCATTCAAGCTATTATTTGTTCCGGAGCC 594
Db 529 TATTTCCTCGGGGAGCTTGAGGTTCCCGTAAGCCATTCAGAGTTACAAATTCCTTTCTGGACCC 588

Qy 595 CACGGACCTTCAAAGTGCAGTGAGCATCCTCGGCTCTGACCCCACAGAGCAATGGGACTTT 654
Db 589 GGGCAAACTTTATCAGGGTCTTGAGTGTCCTCGAACCTCACCACTGGGCACGGGACCTT 648

Qy 655 GACTGGCGTGGGCTACCTTGGAAGAGACCTGAAAGGCCCGCAAGTGCTGCAACTGTAAATCTCAC 714
Db 649 GACTTGTGTGGGCAGAGCTTGAAGGACTTGCAGCCAGCAAGTCCTTAACTGTCAACTGAC 708

Qy 715 TGTGATTCGGGTGTCCCCAAAGCACACTGGAGGTGGTGTATTACCAGGTGTGATTATCAAG 774
Db 709 TGTGGTTCAGCTCCACCTGACAGTATTGGAGAGGA 744

Qy 775 TTTACCGGATTTAGGTTTTTCATTGCCTACTTGGGGCAAAAGTTGGGACTTGGACTAGCAGG 834
Db 745 AGGCCCAGCACTGCCGACCTGGGCCATCATCCTGCTGGGCAGTGGCCTT 792

Qy 835 CACCATGCGTTCTGACGCCGACGTGTACTCTTACAAATACGGCTCGTGCTGCCGCCGTCG 894
Db 793 TTCCTTGCTTGTTGATCCTGATTGTTTTGATTAATAATTCTGTTGCTG 843

Qy 895 TTGTTGTTGGGCTGCAACTGCCGTGTTGTTTCTGCTGTAGAGAAAAGAGGATT 954
Db 844 ---------TTGTGCCTCCAGGAGAGAAGAAGAATCTACTTATCAAATGAAATAAGGAAAT 899

Qy 955 TCGTATTCAATTTCAAAGAAAAICTGAAAAAGAGAGACAAAACAAACAGTGAGACAGA 1014
Db 900 CTGCAAAACATGAGGACAA----------ACAAAGACAGATCCGAGGAGACAAAGTTAAA 945

Qy 1015 AAGTGGAAAATGAAAACTCCGGCTACAAATTCAGATGAACAAAAGACCACAGAAACCGCTTC 1074
Db 946 AAGTGGAAAAGGAAAACTACGGGTTACAGTTGCACAGACTGCATC 1005

Qy 1075 TCTCCCCTCCCAAAATCCTTGTGAATCCAGTCTGAACAAAGACTAGCTGTGTGGCCC 1134
Db 1006 TCTCCCCTAAATCTGCTGAAGTCAGCCTTCCAGAAAAAACGCACAGAGTA-----GCCT 1059

Qy 1135 TCCTCACCAGCCGGGCCTGATCAACGTCCACCAGGCTTCTTT 1194
Db 1060 TCCTTATCAGGAGAACTCAATAAACATCAGCGCCGGTTTCCTT 1119

Qy 1195 TAATCTGGCCAGTGCCTTGAGAAGGTCAGTAATACAACTGTAGTATA 1239
Db 1120 TGACATCGCCAGTCCTCAGAAGGTCAGAAAATGTGACTTTAGTGTA 1164

Query Match   15.2%; Score 188.8; DB 22; Length 401;
Best Local Similarity   99.0%;
Matches 190; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy 115 AGGTTCTGGGTCTGGTAATGAAGTCAATAGAAGGCCCCAGAAATGCAACAGTCCTGAAGGG 174
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 210 AGGTTCTGGGTCTGGTAATGAAGTCAATAGAAGGCCCCAGAAATGCAACAGTCCTGAAGGG 269

Qy 175 CTCCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGGAAGCTTCATCATGTGGGCTCT 234
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 270 CTCCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGGAAGCTTCATCATGTGGGCTCT 329

Qy 235 CAGTGACATGGTGGTGCTTAGGCGTCAGGCCCATGGAGGCGCCCATCATCACCAATGACCGCTT 294
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 330 CAGTGACATGGTGGTGCTTAGGCGTCAGGCCCATGGAGGCGCCCATCATCACCAATGACCGCTT 389

Qy 295 CACCTCTCAGAG 306
       ||||||||||||
Db 390 CACCTCTCAGAG 401

```
Qy    241 AATCTTGTAGTCGCTGAGAATGAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACC 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    524 CGGCTCCCGGATATTCCTGGGGAGCTCGGCTCCTCGGTCTCCTGGTCTCAGCCATTCTATTTATTT 583

Qy    301 CGGCTCCCGGATATTCCTGGGGAGCTCGGCTCCTCGGTCTCCTGGTCTCAGCCATTCTATTTATTT 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    584 GTTCCGGAGCCCAGCGAGCCTTCAAAGTGCAGTGAGCAATCCTGGCTCTGACCCCACAGAGC 643

Qy    361 GTTCCGGAGCCCAGCGAGCCTTCAAAGTGCAGTGAGCAATCCTGGCTCTGACCCCACAGAGC 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    644 AATGGGACTTTGACTTGCCTGGCTACCTGGGAACAGAGACCCGCAAGTGTGCAACT 703

Qy    421 AATGGGACTTTGACTTGCCTGGCTACCTGGGAACAGAGACCCGCAAGTGTGCAACT 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    704 GTAAATCTCACTGTGATTCGGTGTCCCAAGACACTGGAGGTGGTATTAATTCCAGGT 763

Qy    481 GTAAATCTCACTGTGATTCGGTGTCCCAAGACACTGGAGGTGGTATTAATTCCAGGT 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    764 GTATTATCAAGTTTACCGAGTTTAGGTTTTTCAATGCCTACTTGGGGCGAAAGTTGGGACTT 823

Qy    541 GTATTATCAAGTTTACCGAGTTTAGGTTTTTCAATGCCTACTTGGGGCGAAAGTTGGGACTT 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    824 GGACTAGCAGGCAGCCACCATGCTGTTCTGACGCGCGACGTGTACTCTTACAAATACGGCTGCTGCTGCTGC 883

Qy    601 GGACTAGCAGGCAGCCACCATGCTGTTCTGACGCGCGACGTGTACTCTTACAAATACGGCTGCTGCTGC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    884 TGCGCGCCGTCGTTGTTGTGGGCTGCTGCCCGTTGTTGTTCTTGTCTGTAGAAGA 943

Qy    661 TGCGCGCCGTCGTTGTTGTGGGCTGCTGCCCGTTGTTGTTCTTGTCTGTAGAAGA 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    944 AAAAGAGGATTTCGTATTC- 962

Qy    721 AAAAGAGGTGACAATCTTCTGACAATCTTCCCACAATCCCAGAGATTAGGTGCATGCATC 780
          ||||||     |    |||
Db    963 --------AATTTCAAAGAGAAATCTGAAAAGAAGAAGAAAGAAAGAAAGAAACTGAGACAG 1013

Qy    781 AATGGCGTCATTTTGTGTATACTCATTTATNTGTCTGGCCTTCTTGAGCCTCTCACTGTGAGA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1014 AAAGTGGGAAATGAAAAACTCCGGCTACTACAAATTCAGATGAATGAACCACAGAAAACCGCTT 1073

Qy    841 CTGATCCATACCGTGGCACGAAATCTGAAAAGAAGAAGAAGAAAGAAAGAAAGAAACTGAGACAG 900
          |   ||||||||||||||||||||||||||||||||||
Db    901 AAAGTGGGAAATGAAAAACTCCGGCTACTACAAATTCAGATGAATGAACCACAGAAAACCGCTT 960

Qy    901 AAAGTGGGAAATGAAAAACTCCGGCTACTACAAATTCAGATGAATGAACCACAGAAAACCGCTT 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 CTCTCCCTCCAAAATCCTGTGAATCCAGTGATCCTGAACAAGAAGAACAGTAGCTGTGTGGCC 1020

Qy    963 ---------------- 962

Qy   1014 AAAGTGGGAAATGAAAAACTCCGGCTACTACAAATTCAGATGAATGAACCACAGAAAACCGCTT 1073
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTCTCCACCAGAGAAGAACAATAGACCT 1046

Qy   1074 CTCTCCCTCCAAAATCCTGTGAATCCAGTGATCCTGAACAAGAAGAACAGTAGCTGTGTGGCC 1133
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTCTCCACCAGAGAAGAACAATAGACCT 1046

Qy   1134 CTCCTCACCAGCGCGGCCTGATCAACGT 1159
          ||||||||||||   |    ||  |  ||
Db   1021 CTCTCCACCAGAGAAGAACAATAGACCT 1046
```

RESULT 5

AAS92356
ID   AAS92356 standard; cDNA; 1392 BP.
XX
AC   AAS92356;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #28160.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT, Liu C, Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG28169.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID NO 28160; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (I) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1392 BP; 338 A; 361 C; 341 G; 352 T; 0 other;

Query Match   62.7%; Score 777.6; DB 23; Length 1392;
Best Local Similarity   87.2%;
Matches 912; Conservative   0; Mismatches 24; Indels 110; Gaps   1;

Qy   224 ARTGTGGGCCTCTCAGTGAACATGGTGGTGGTCGTGCTTAGGGTGCAGGCCCATCANCACC 283
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ARTGTGGGCCTCTCAGTGACATGGTGGTGGTCGTGCTTAAGCGTCAGGCGCCCATCANCACC 60

Qy   284 AANTGACCGGCTTCACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATC 343
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AANTGACCGGCTTCACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATC 120

Qy   344 ATCCACAAATGTGGAGCCCAGTGATTCGGGGAACATCAGATGCAGCCTTCAGAACAGTCGC 403
          ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db   121 ATCCCACAAATGTGGAGCCCAGTGAATTCGGGGAACATCAGATCAGATGCAGCCTTCAGAACAGTCGC 180

Qy   404 CTGCATGGGATCGTGCTTPACCTTACCGTCCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTT 463
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 CTGCATGGGATCGTGCTTACCTTACCGTCCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTT 240

Qy   464 AATCTTGTAGTGTCGCTCAGAATGAACCTTGTGAAGTTACTTGTGTCTACCCTCACACTGGACC 523
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 ...

```
Qy    646  TGGGACTTTGACTTGCGTGCGTGGCTACCTGGAAGAGAGCCTGAAGAGCCCGCAAGTCTGCAAACTGT  705
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    574  TGGGACTTTGACTTGCGTGCGTGGCTACCTGGAAGAGAGCCTGAAGAGCCCGCAAGTCTGCAAACTGT  633

Qy    706  AAATCTCACTGTGATTCGGTGTCCCCAAGACACTGGAGGTGGTGATTAATAATATTCCAGGTGT  765
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    634  AAATCTCACTGTGATTCGGTGTCCCCAAGACACTGGAGGTGGTGATTAATAATATTCCAGGTGT  693

Qy    766  ATTATCAAGTTACCGGAGTTTAGGTTTTTCATTCCCTACTCTTGGGCAAAGTTTGGACTTGG  825
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    694  ATTATCAAGTTACCGGAGTTTTTCATTCCCTACTCTTGGGCAAAGTTTGGACTTGG  753

Qy    826  ACTAGCAGGCACCATGCTTCTGACGCGTGACGTGACGTTGACGCCGACGTGTACTCGTGCTGCTG  885
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    754  ACTAGCAGGCACCATGCTTCTGACGCGTCTTTACAATACGCCTTGGCTGCTG  813

Qy    886  CCGCCGTCGTTGTGTTGGGCGTCGCCAACTGGCTGCTGCTGGTTGTTTCTGCTGTAGGAGAAA  945
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    814  CCGCCGTCGTTGTGTTGGGCGTCGCAACTGGCTGCTGCTGGTTGTTTCTGCTGTAGGAGAAA  873

Qy    946  AAGAGGGATTTCGTATTCAATTTCAAAAAGAAAATCGAAAAAGAAGAGAAGAGACAAAACAAAGAAAC 1005
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    874  AAGAGGGATTTCGTATTCAATTTCAAAAAGAAAATCTGAAAAAAGAGAGAGACAAAACAAAGAAAC  933

Qy   1006  TGAGACAGAAAAGTGGAAAATGAAAAACTCCGGCTACAATTCAGATGAACAAAAGACCACAGA 1065
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    934  TGAGACAGAAAAGTGGAAAATGAAAACTCCGGCTACAATTCAGATGAACAAAAGACCACACAGA  993

Qy   1066  AACCGCTTCTCTCCCCCCCAAAATCCTGTGATTCCAGTGATTCGGACAAAGAAGAAACAGTAG 1125
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    994  CACCGCTTCTCTCCCCCCAAAATCCTGTGAATCCAGTGATTCCGGACAAAGAAGAAACAGTAG 1053

Qy   1126  CTGTGTGCCCCCTCACCACGGGGGGGCGTCCACCCAGGCCAGCCAAGTCATCCCACA 1185
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1054  CTGTGGCCCCTCCACCACGCGGGCGTAATCAACGTCACCACCCAGGCCAAGTCATCCCACA 1113

Qy   1186  GGCTTCTTTTAAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG 1240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1114  GGCTTCTTTTAAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG 1168
```

RESULT 4

ABK13031

ID   ABK13031 standard; cDNA; 1139 BP.
XX
AC   ABK13031;
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h4.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; allergy; nephropathy; skin disorder;
KW   toxic shock syndrome; graft versus host disease; autoimmune disease;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1134
FT   /*tag= a
FT   /product= "B7-like protein, B7-L_h4"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20119.
XX
PR   28-JUN-2000; 2000US-2145122P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis
XX
PS   Claim 1; Fig 4; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility) and
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions, various
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia, and
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h4.
XX
SQ   Sequence 1139 BP; 290 A; 300 C; 283 G; 266 T; 0 other;

```
Query Match             85.8%;  Score 1064;  DB 24;  Length 1139;
Best Local Similarity   97.2%;
Matches 1102;  Conservative   0;  Mismatches  10;  Indels  22;  Gaps   1;

Qy   106  AGAACGTGTAGGTTCGTGGGTCTGGTGTAATGAATGTCATAGAAGGTCATAGAAGGTCAACAGT  165
          || |||||| | |||||||||||||||| ||||||||||||||||||||| |||||| |
Db    27  AGACCCACCCGGTTCGTGGGTCTGGTPAATGAAGTCATAGAAGGCCCCCAGAAATGCAACAGT  86

Qy   166  CCTGAAGGGCTCCCAGGCTCCGCTTCAACTGCACCGTGCTCCCAGGGGCTGGAAGCTCATCAT  225
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    87  CCTGAAGGGCTCGCCAGGCTCGCTTCAACTGCACCGTGCTCCCAGGGCTGGAAGCTCATCAT  146

Qy   226  GTGGGCTCTCAGTGACATGGTGGTGGTGCTAACGGTCAGGCCGACCCATCATCACCACCAA  285
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   147  GTGGGCTCTCAGTGACATGGTGGTGGTGCTAACGGTCAGGCCGACCCATCATCACCACCAA  206

Qy   286  TGACCGCTTCACCTCTCAGAGGTACGACCAGGGGCGGGAACTTCACCTCGGAGATGATCAT  345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   286  ...
```

```
Db   891 CGTATTCAATTTCAAAAGAATCTGAAAAGAGAGAAGAAGAAAACTGAGAGACAGAA 950
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1016 AGTGGAAATGAAAACTCCGGCGTACAATTCAGATGAACAAAAGAGAACAAGAAAACTGAGAGACAGAA 1075

Db   951 AGTGGAAATGAAAACTCCGGCGTACAATTCAGATGAACAAAAGAGAACACCGCTTCT 1010
         ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Qy  1076 CTCCCTCCCAAATCCTGTGAATCCACTGAGATCCTGAACAAAGAAAAGAGGTACACTGTGCCCT 1135

Db  1011 CTCCCTCCCAAATCCTGTGAATCCACTGAGATCCTGAACAAAGAAAAGAGTACCTGTGCCCT 1070
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1136 CCTCACCAGGGGCTGTGATCAACGTCACCACCAGGGCCAGCAAGTCATCCACCACAGGCCTTCTTTT 1195

Db  1071 CCTCACCAGGGGGTCTGATCAACGTCACCACCAGGGCAAGTCATCCACCACAGGCCTTCTTTT 1130
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1196 AATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTTAGTAGTAG 1240

Db  1131 AATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTTAGTAGTAG 1175
```

RESULT 3

ABK13029

```
ID   ABK13029  standard; cDNA; 1168 BP.
XX
AC   ABK13029;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h2.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FH
FT   CDS            8..1168
FT                  /tag= a
FT                  /product= "B7-like protein, B7-L_h2"
XX
PN   WO200200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2000; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75541.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PT   Claim 1; Fig 2; 135pp; English.
XX
PS   The invention relates to an isolated B7-like (B7-L) polypeptide (I),
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility),
CC   miscarriage, preterm labour and delivery and endometriosis) and
```

```
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenia
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h2.
XX
SQ   Sequence 1168 BP; 299 A; 302 C; 288 G; 279 T; 0 other;

Query Match             90.2%;  Score 1119;  DB 24;  Length 1168;
Best Local Similarity   99.1%;
Matches 1125;  Conservative   0;  Mismatches   10;  Indels   0;  Gaps   0;

Qy  106 AGAAGCTGTAGGTTCTCGGGTCTCGGTAATGAAGTCATAGAAGGCCCCCAGAATGCAACAGT 165
        ||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||  |||
Db   34 AGACCCACCCGGTTCTGGGTCTCGGTAATGAAGTCATAGAAGGCCCCCAAAATGCAAGAGT 93

Qy  166 CCTGAAGGGCTCCCAGGGCTCGCTTCAACTGCACCGTCTCCCAGGGGGTGGAAGCTCATCAT 225
        |||||| |||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   94 CCTGAAAGGGCTCCCAGGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCAT 153

Qy  226 GTGGGCTCTCAGTGACATGTGGTGCTAAGGCGTCAGGCCCCATGGAGCCCCATCATCACCAA 285
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  154 GTGGGCTCTCAGTGACATGTGGTGCTAAGGCGTCAGGCCCCATGGAGCCCCATCATCACCAA 213

Qy  286 TGACCGCTTCACCTCTCAGAGGTACGACCAGGCGGGAACTTCACCTCGGAGATGATCAT 345
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  214 TGACCGCTTCACCTCTCAGAGGTACGACCAGGCGGGAACTTCACCTCGGAGATGATCAT 273

Qy  346 CCACAAATGTGGAGCCCAGTGATTCGGGGGAACATCAGATGCAGCCTCCAGAACAGTCGCCT 405
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  274 CCACAAATGTGGAGCCCAGTGATTCGGGGGAACATCAGATGCAGCCTCCAGAACAGTCGCCT 333

Qy  406 GCATGGATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAA 465
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  334 GCATGGATCTGCTTACCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAA 393

Qy  466 TCTTGTAGTCCGTCGAGAATGAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACCCG 525
        |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db  394 TCTTGTAGTCGGTCGAGAATGAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACCCG 453

Qy  526 GCTCCCGGATATTTCCTGGGAGCTCGGTCTCCTGGTCAGCCGCCATTCAAGCTATTATTTGT 585
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  454 GCTCCCGGATATTTCCTGGGAGCTCGGTCTCCTGGTCAGCCGCCATTCAAGCTATTATTTGT 513

Qy  586 TCCGGAGCCCAGCGACCTTCAAAGTCCAGTGAGCATCCTGGCTCTGGACCCACACAGAGCAA 645
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  514 TCCGGAGCCCAGCGACCTTCAAAGTCCAGTGAGCATCCTGGCTCTGGACCCACACAGAGCAA 573
```

```
ID   ABK13028 standard; cDNA; 1175 BP.
XX
AC   ABK13028;
XX
DT   23-APR-2002 (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h1.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; antipsoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; allergy; nephropathy; skin disorder; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            27..1175
FT                  /*tag= a
FT                  /product= "B7-like protein, B7-L_h1"
XX
PN   WO020200710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
XX
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75540.
XX
PT   New B7-like polypeptides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 1; Fig 1; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (I) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, pretern labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (I) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   in allograft transplantation, graft versus host disease, T-cell
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h1.
SQ   Sequence 1175 BP; 295 A; 307 C; 286 G; 287 T; 0 other;

Query Match      90.6%;  Score 1123.4;  DB 24;  Length 1175;
Best Local Similarity   99.9%;   Pred. No. 1.6e-300;
Matches 1124;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy   116 GGTTCTGGGTCTGGTGGTTGGTGAATGAAGTCATAGAAGTCATAGAGAATGCAACAGTCCTGAAGGGC 175
         ||||||||||||||                |||||||||||||||||||||||||||||||||||||||
Db    51 GGTTCTGGGTCTGGTAATGAAGTCATAGAAGTCATAGAGAATGCAACAGTCCTGAAGGGC 110

Qy   176 TCCCAGGCCTCGGCTTCAACTGCACCGTTCTCCCAGGCCTGGAGGCTCATCATGTGGGCTCTC 235
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   111 TCCCAGGCCTCGGCTTCAACTGCACCGTTCTCCCAGGCCTGGAGGCTCATCATGTGGGCTCTC 170

Qy   236 AGTGACACATGGTGGTGCTGTCTGCTAAGCGTGCAGCCCCATGGAGGCCCATCATCACCAATGACCGCTTC 295
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   171 AGTGACACATGGTGGTGCTGTCTGCTAAGCGTGCAGCCCCATGGAGGCCCATCATCACCAATGACCGCTTC 230

Qy   296 ACCTCTCAGAGGTACGACCAGGCGGGAACTTCACCTCGGAGAGTGATCATCCACAATGTG 355
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   231 ACCTCTCAGAGGTACGACCAGGCGGGAACTTCACCTCGGAGAGTGATCATCCACAATGTG 290

Qy   356 GAGCCCAGTGAATTCGGGGAACATCAGATGCAGCCTTCCAGAACAGCAGTCGCCGTCGCATGGATCT 415
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   291 GAGCCCAGTGAATTCGGGGGAACATCAGATGCAGCCTTCCAGAACACAGTCGCCGTCGCATGGATCT 350

Qy   416 GCTTACCTTACCGTCCAAGTTATGGGGAGAGCTGTTCAATTCCCAGTGTGTTAATCTTGTAGTC 475
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   351 GCTTACCTTACCGTCCAAGTTATGGGGAGAGCTGTTCAATTCCCAGTGTGTTAATCTTGTAGTC 410

Qy   476 GCTGAGAATGAACCTTGTGAAGTTACTTGTCTACCCTCCACACTGGACCGGGGTCCCGGAT 535
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   411 GCTGAGAATGAACCTTGTGAAGTTACTTGTCTACCCTCCACACTGGACCGGGGTCCCGGAT 470

Qy   536 ATTTCCTGGGGAGCTCGTGTCTCCTGGTCAGCCAATCAAGCTATTATTTTGTTGTCAAGT 595
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   471 ATTTCCTGGGGAGCTCGGTGTCTCCTGGTCAGCCAATCAAGCTATTATTTTGTTGTCAAGT 530

Qy   596 AGCGGACCTTCAAAGTGCAGTGAAGCATCCTGGCTCTGACCCACAGAGGCCAATGGGACTTTG 655
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   531 AGCGGACCTTCAAAGTGCAGTGAAGCATCCTGGCTCTGACCCACAGAGGCCAATGGGACTTTG 590

Qy   656 ACTTGCCGTGCTACCTGGAAGAGCCTGAAGGGCCCGCAAGTCTGCAACTGTAAATCTCACT 715
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   591 ACTTGCCGTGCTACCTGGAAGAGCCTGAAGGGCCCGCAAGTCTGCAACTGTAAATCTCACT 650

Qy   716 GTGATTCGGTGTCCCCAGAACACTGGAGTGGGTTATTAATATTCCAGGTGTTATTATCAAGT 775
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   651 GTGATTCGGTGTCCCCAGAACACTGGAGTGGGTTATTAATATTCCAGGTGTTATTATCAAGT 710

Qy   776 TTACCGAGTTTAGGTTTTTCATTGCCCTACTTGGGCAAGTTGGACTTAGCAGGC 835
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   711 TTACCGAGTTTAGGTTTTTCATTGCCTACTTGGGCAAAGTTGGACTTAGCAGGC 770

Qy   836 ACCATGCTTCTGACGGCCGACGTGTACTCTTACAATACGCTGCTGGTCGCGTCGT 895
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   771 ACCATGCTTCTGACGGCCGACGTGTACTCTTACAATACGCTGCTGGTCGCGTCGT 830

Qy   896 TGTTTGTGCTGCTCGCTGTGTCCGTTGTTGTTCTGCTGAGAAGAAAAAGAGGAGATTT 955
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   831 TGTTTGTGCTGCAACTGCTGCTCGCCGTTGTTGTTCTGCTGTGAGAAGAAAAAGAGGAGATTT 890

Qy   956 CGTATTCAATTCAAAAAGAAATCTGAAAAAGAGACAAGACAAAACAAACTGAACACAGAA 1015
```

```
PR   28-JUN-2000; 2000US-214512P.
PR   28-NOV-2000; 2000US-0729264.
XX
PA   (AMGE-) AMGEN INC.
XX
PI   Welcher AA, Sarmiento UM, Schultz HJ, Chute HT;
XX
DR   WPI; 2002-130881/17.
DR   P-PSDB; AAU75542.
XX
PT   New B7-like polypeptides, polynucleotides and their modulators, useful
PT   for diagnosing, preventing and treating reproductive, immune and
PT   proliferative disorders, e.g. cancer and arteriosclerosis -
XX
PS   Claim 1; Fig 3; 135pp; English.
XX
CC   The invention relates to an isolated B7-like (B7-L) polypeptide (I).
CC   The polypeptide, polynucleotide encoding it and antibody against (II) are
CC   useful for treating B7-like polypeptide-related disease, disorders or
CC   conditions including reproductive disorders (e.g. infertility,
CC   miscarriage, preterm labour and delivery and endometriosis) and
CC   proliferative disorders. Antibodies, soluble proteins comprising
CC   extracellular domains and other regulators of B7-L polypeptides are
CC   useful for enhancing the immune response to tumours. (i) plays a role in
CC   growth and maintenance of cancer cells based on the observation of
CC   seminal vesicle hyperplasia in transgenic mice overexpressing B7-L
CC   polypeptide. Hence modulators of (I) are useful for the treatment of
CC   cancer including seminal vesicle cancer, lung, brain, breast, ovarian,
CC   testicular cancer and cancers of haematopoietic system. B7-L polypeptide
CC   in allograft transplantation, graft versus host disease, T-cell
CC   pathway can be manipulated to regulate cytotoxic T-lymphocyte response
CC   dependent B-cell mediated diseases and autoimmune diseases. B7-L
CC   molecules are useful for alleviating the symptoms associated with
CC   diseases involving chronic immune cell dysfunction or to treat
CC   arthritis, multiple sclerosis, diabetes, immune thrombocytopenic
CC   autoimmune diseases such as systemic lupus erythematosus, rheumatoid
CC   purpura and psoriasis, chronic inflammatory disease such as
CC   inflammatory bowel disease (Crohn's disease and ulcerative colitis),
CC   Grave's disease, Hashimoto's thyroiditis and diabetes mellitus. They
CC   are also useful as immunosuppressive agents for bone marrow and organ
CC   transplantation or to prolong graft survival. B7-L molecules are also
CC   useful for diagnosis and treatment of diseases involving abnormal cell
CC   proliferation, including arteriosclerosis and vascular restenosis.
CC   Antagonists of B7-L polypeptides are useful for alleviation of toxic
CC   shock syndrome or allosensitisation due to blood transfusions, and for
CC   treatment of allergy, asthma and hypersensitivity reactions,
CC   nephropathies (e.g. glomerulonephritis), skin disorders (pemphigus and
CC   pemphigoid), endocrinopathies (Grave's disease), various pneumopathies
CC   (extrinsic alveolitis), vasculopathies, coeliac disease, anaemia,
CC   thrombocytopenias, Guillain-Barre syndrome and myasthenia gravis, and
CC   lymphoproliferative disorders such as multiple myeloma. The present
CC   sequence represents the coding sequence of human B7-L_h3.
XX
SQ   Sequence 1240 BP; 318 A; 319 C; 305 G; 298 T; 0 other;

Query Match             100.0%;  Score 1240;  DB 24;  Length 1240;
Best Local Similarity   100.0%;
Matches 1240;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy         1 AGGTGTGAGTCCACCCACACAGTGTGGATCAGTTCTAGGCTCGCCAATAACAAAGCACCAT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1 AGGTGTGAGTCCACCCACACAGTGTGGATCAGTTCTAGGCTCGCCAATAACAAAGCACCAT 60

Qy        61 AACCTGGTCGGCTTAGAACAAATGGAAAGGCCATTTGCTCACGGTTCCAGAAGCTGTAGGTTC 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61 AACCTGGTCGGCTTAGAACAAATGGAAAGGCCATTTGCTCACGGTTCCAGAAGCTGTAGGTTC 120

Qy       121 TGGGTCTCTGGTAAATGAAGTCATTAGAAGGCCCCCAGAATCCAAACAGTCCTGAAGGGCTCCCA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 TGGGTCTCTGGTAAATGAAGTCATTAGAAGGCCCCCAGAATCCAAACAGTCCTGAAGGGCTCCCA 180

Qy       181 GGCTCGCTTCAACTGCACCGTCTCCCAGGGCGTCTGGAAGCTGCATCTCATCTATGTGGGCGCTCAGTGA 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 GGCTCGCTTCAACTGCACCGTCTCCCAGGGCGTCTGGAAGCTGCATCTCATCTATGTGGGCGCTCAGTGA 240

Db       181 GGCTGCTTCAACTGCACCGTCTCCCAGGGCGTGGGAAGCTCATCATGCACTTC 240
Qy       241 CATGGTTGGTGCTTAAGCGTCAGGCCCATCACCAATGACCGCTTCACCTTC 300
             |||||||||||||||||||||||||||||||||||||||||||||
Db       241 CATGGTTGGTGCTTAAGCGTCAGGCCCATCACCAATGACCGCTTCACCTTC 300

Qy       301 TCAGAGGTACGACCAGGCGGGAACTTCACCTCGGAGATGAATCATCACAAATGTGGGAGCC 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 TCAGAGGTACGACCAGGCGGGAACTTCACCTCGGAGATGAATCATCACAAATGTGGGAGCC 360

Qy       361 CAGTGAATTCGGGGAACATCAGATGCAGCCTTCCAGAACAGTCGCGCTGGATCTGCTTA 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 CAGTGAATTCGGGGAACATCAGATGCAGCCTTCCAGAACAGTCGCGCTGGATCTGCTTA 420

Qy       421 CCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCACTGGTTTGTAGTTCGGTGA 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 CCTTACCGTCCAAGTTATGGGAGAGCTGTTCATTCCCACTGGTTTGTAGTTCGGTGA 480

Qy       481 GAATGAAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACCGGGTCCCCGGGATATTTC 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 GAATGAAACCTTGTGAAGTTACTTGTCTACCCTCACACTGGACCGGGTCCCCGGGATATTTC 540

Qy       541 CTGGGAGCTCGGTCTCCAGGCATTCAAGCTATATATTGTTCGGGAGCCCAGCGA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 CTGGGAGCTCGGTCTCCAGGCATTCAAGCTATATATTGTTCGGGAGCCCAGCGA 600

Qy       601 CCTTCAAAAGTGCAGTGAGCATTCCTGGCCTCTGACCCCACAGAGCAAATGGGACTTTGACTTG 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 CCTTCAAAAGTGCAGTGAGCATTCCTGGCCTCTGACCCCACAGAGCAAATGGGACTTTGACTTG 660

Qy       661 CGTGGCTACCTGGAAGAGCCTGAAAGGCCCGAGTCTGCAACTGTCAAATCTCACTGTGAT 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 CGTGGCTACCTGGAAGAGCCTGAAAGGCCCGAGTCTGCAACTGTCAAATCTCACTGTGAT 720

Qy       721 TCGGTGTCCCAAGACACTGGAGGTGGTTATTAATATTCCAGGTGTATTATATCAAGTTACC 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       721 TCGGTGTCCCAAGACACTGGAGGTGGTTATTAATATTCCAGGTGTATTATATCAAGTTACC 780

Qy       781 GAGTTTAGGTTTTCATTGCCTACTTGGGCCAAAGTTGGGACTTGGACTAGCAGGCACCAT 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       781 GAGTTTAGGTTTTCATTGCCTACTTGGGCCAAAGTTGGGACTTGGACTAGCAGGCACCAT 840

Qy       841 GCTTCTGACGCCGACGTGTACTTCTTACAATACGGCTGTCGTGCGCCGCGTTGCTTG 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       841 GCTTCTGACGCCGACGTGTACTTCTTACAATACGGCTGTCGTGCGCCGCGTTGCTTG 900

Qy       901 TGGGCTGCAACTGCTGCTGCGGTGTGTTGTTTCTGCTGTGTAGAAGGAAAGAGAGGATTCGTAT 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       901 TGGGCTGCAACTGCTGCTGCGGTGTGTTGTTTCTGCTGTGTAGAAGGAAAGAGAGGATTTCGTAT 960

Qy       961 TCAATTTCAAAAGAAACTCTGAAAAAAAAGACAAAAACAAAAGAAACTGAGACAGAAAGTGG 1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       961 TCAATTTCAAAAGAAACTCTGAAAAAAAAGACAAAAACAAAAGAAACTGAGACAGAAAGTGG 1020

Qy      1021 AAATGAAAACTCCGGCTACAAATTCAGATGAACAAAAGACCACAGAAAACGCTTCTCTCCC 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1021 AAATGAAAACTCCGGCTACAAATTCAGATGAACAAAAGACCACAGAAAACGCTTCTCTCCC 1080

Qy      1081 TCCCAAAATCCTGTGAATCCAGTGAATCCTGAACAAAAGAAAACAGTAGCTGTGTGGGCCCAT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1081 TCCCAAAATCCTGTGAATCCAGTGAATCCTGAACAAAAGAAAACAGTAGCTGTGTGGGCCCTCCA 1140

Qy      1141 CCAGCGGGGCCTGATCAAGTCACCCAGCCACAGCACGCTTCTCTTTAATCT 1200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1141 CCAGCGGGGCCTGATCAAGTCACCCAGCCACAGCACGCTTCTCTTTAATCT 1200

Qy      1201 TGGGTCTCGGTAATGAAGTCAGTAATACAAACTTAGTATAG 1240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1201 GGCCAGTCCTGAGAAGGTCAGTAATACAAACTTAGTATAG 1240
```

RESULT 2
ABK13028

```
OM nucleic - nucleic search, using sw model

                    GenCore version 5.1.4_p5_4578
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:             April 28, 2003, 23:25:03 ; Search time 300.05 Seconds
                                                (without alignments)
                                                9306.701 Million cell updates/sec

Title:              US-09-729-264-5
Perfect score:      1240
Sequence:           1 aggtgtgagtccagccaaca.........gtaatacaactgtagtatag 1240
Scoring table:      IDENTITY_NUC
                    GapOP 10.0 , Gapext 1.0
Searched:           2185239 seqs, 112599159 residues
Total number of hits satisfying chosen parameters:  4370478

Post-processing:    Minimum Match 0%
                    Maximum Match 100%
                    Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :          N_Geneseq_101002:*
            1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
            2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
            3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
            4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
            5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
            6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
            7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
            8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
            9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
           10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
           11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
           12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
           13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
           14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
           15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
           16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
           17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
           18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
           19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
           20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
           21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
           22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
           23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
           24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1240	100.0	1240	24	ABK13030	DNA encoding human
2	1123.4	90.6	1175	24	ABK13028	DNA encoding human
3	1119	90.2	1168	24	ABK13029	DNA encoding human
4	1064	85.8	1139	24	ABK13031	DNA encoding human
5	777.6	62.7	1392	23	AAS92356	DNA encoding novel
6	380.6	30.7	1195	24	ABK13032	DNA encoding mouse
7	188.8	15.2	401	22	ABK13033	Probe #5268 used t
8	185.6	15.0	754	24	AAI36582	DNA encoding mouse
9	185.6	15.0	895	24	ABK13033	DNA encoding mouse

c	10	53.4	4.3	277	21	ABK10239	Trinucleotide repe
c	11	52.8	4.3	7029	23	ABL29756	Drosophila melanog
c	12	51.8	4.2	389	21	ABK10248	Tri-nucleotide rep
c	13	51.8	4.2	397	20	AAX898891	Spinocerebellar at
c	14	51.8	4.2	403	21	ABK10240	Trinucleotide repe
c	15	49.6	4.0	742	22	AA195221	Human neuroblastom
c	16	49.4	4.0	46954	23	ABL16830	Drosophila melanog
c	17	48.4	3.9	3217	10	AAN91578.	Rat androgen recep
c	18	48.4	3.9	3217	12	AAQ12002	Full-length rat an
c	19	48.4	3.9	4180	21	AAN91773	Rat androgen recep
c	20	47.6	3.8	2733	20	AAX84442	Mouse brain CNG-1
c	21	47.4	3.8	3894	23	ABL03353	Drosophila melanog
c	22	47.4	3.8	22341	23	ABL03352	Drosophila melanog
c	23	46.6	3.8	300	20	AAX98588	Human cancer cell
c	24	46.6	3.8	759	20	AAX98974	Human validated ca
c	25	46.6	3.8	2333	22	AAF75339	Human TGF-beta rec
c	26	46.6	3.8	14208	22	AAK78411	Human immune/haema
c	27	46.4	3.7	1983	23	ABL24483	Drosophila melanog
c	28	46.4	3.7	4044	23	ABL24482	Drosophila melanog
c	29	46.4	3.7	14152	22	AAS36073	Human cardiovascul
c	30	46.4	3.7	14152	22	AAK82884	Human immune/haema
c	31	46.2	3.7	381	22	AAV44466	Human prostate exp
c	32	46	3.7	513	23	ABV47976	Human prostate exp
c	33	45.8	3.7	405	23	ABV56338	Human prostate exp
c	34	45.8	3.7	2682	23	ABL16096	Drosophila melanog
c	35	45.8	3.7	8821	23	ABL16097	Drosophila melanog
c	36	45.6	3.7	1509	23	ABL24571	Drosophila melanog
c	37	45.6	3.7	3884	23	ABL24570	Drosophila melanog
c	38	45.4	3.7	578	22	AAS30941	Human cDNA encodin
c	39	45.4	3.7	3042	23	ABL28445	Human cDNA encodin
c	40	45.4	3.7	3135	23	ABL25653	Drosophila melanog
c	41	45.4	3.7	5215	23	ABL28444	Drosophila melanog
c	42	45.4	3.7	5447	23	ABL25652	Drosophila melanog
c	43	45.2	3.7	2134	22	AAF93812	Human cDNA encodin
c	44	45.2	3.6	5397	23	ABL29757	Drosophila melanog
c	45	45.2	3.6	462	23	ABL26941	Drosophila melanog

ALIGNMENTS

```
RESULT 1
ABK13030
ID   ABK13030 standard; cDNA; 1240 BP.
XX
AC   ABK13030;
XX
DT   23-APR-2002  (first entry)
XX
DE   DNA encoding human B7-like protein, B7-L_h3.
XX
KW   Human; B7-like protein; B7-L; antiinfertility; gynaecological;
KW   antitumour; cytostatic; immunosuppressive; antiarthritic; antirheumatic;
KW   antiinflammatory; dermatological; psoriatic; neuroprotective;
KW   antidiabetic; haemostatic; antithyroid; antiulcer; antiallergic;
KW   antiasthmatic; nephrotropic; antibacterial; virucide; tumour; cancer;
KW   reproductive disorder; graft versus host disease; autoimmune disease;
KW   toxic shock syndrome; allergy; nephropathy; skin disorder;
KW   endocrinopathy; lymphoproliferative disorder; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             80..1240
FT                   /*tag= a
FT                   /product= "B7-like protein, B7-L_h3"
XX
PN   WO20020710-A2.
XX
PD   03-JAN-2002.
XX
PF   28-JUN-2001; 2001WO-US20719.
```

```
Qy       1 AGGTGTGAGTCCAGCCAACAGTGTGGATCAGTTCCTAGGCTGCCATAACAAAGCACCAT 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  159444 AGGTGTGAGTCCAGCCAACAGTGTGGATCAGTTCCTAGGCTGCCATAACAAAGCACCAT 159503

Qy      61 AACCTGGTGGCTTAGAACAATGGAAAGGCATTTGCTCACGGTTCCAGAAGCTGTAGGTT 119
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  159504 AACCTGGTGGCTTAGAACAATGGAAAGGCATTTGCTCACGGTTCCAGAAGCTGTAGGTT 159562
```

Search completed: April 29, 2003, 14:51:53
Job time : 4716.78 secs

```
RESULT 15
AF064860
LOCUS       AF064860              170121 bp    DNA    linear   PRI 05-MAR-2002
DEFINITION  Homo sapiens chromosome 21 clone PAC 70I24 map 21q22.3, complete
            sequence.
ACCESSION   AF064860
VERSION     AF064860.2  GI:18958624
KEYWORDS    HTG; HTGS_DRAFT.
SOURCE      Homo sapiens
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 170121)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.-S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.-K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M.B., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Kawasaki,K., Asakawa,S.,
            Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Haaf,T., Wehrmeyer,S.,
            Borzym,K., Gardiner,K., Nizetic,D., Francis,F., Lehrach,H.,
            Reinhardt,R. and Yaspo,M.Laure.
  TITLE     The DNA sequence of human chromosome 21
  JOURNAL   Nature 405 (6784), 311-319 (2000)
   MEDLINE  20289799
   PUBMED   10830953
REFERENCE   2  (bases 1 to 170121)
  AUTHORS   Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
            Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAY-1998) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 170121)
  AUTHORS   Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
            Weber,J., Schattevoy,R., Yaspo,M.-L., Rosenthal,A., Yaspo,M.-L. and
            Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-FEB-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 170121)
  AUTHORS   Rump,A., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
            Weber,J., Schattevoy,R., Yaspo,M.-L. and Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
COMMENT     On Feb 27, 2002 this sequence version replaced gi:171153.
FEATURES             Location/Qualifiers
     source          1..170121
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="PAC 70I24"
BASE COUNT    45184 a  36756 c  37509 g  50672 t
ORIGIN
```

Qy   1 AGGTGTGAGTCCAGCCAACACAGTGTGGANTCAGTGTTCCTTAGGTGTGCCATAACAAAACACCAT 60
          ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db 18694 AGGTGTGAGTCCAGCCAACACAGTGTGGATCAGTGTTCCTTAGGTGTGCCATAACAAAGCACCAT 18753

Qy   61 AACCTGGTGGCTTAGAACAATGGAAAGGCATTTGCTCACGGTTCCAGAGGTTCCGAGAAGCTGTTAGGTT 119
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db 18754 AACCTGGTGGCTTAGAACAATGGAAAGGCATTTGCTCACGGTTCCAGAGGTTCCGAGAAGCTGT.AGGTT 18812

```
repeat_region   /note="GC score = 10.60 (249bp); Region: GC content"
                /evidence=not_experimental
        15369..16268
                /evidence=not_experimental
                /rpt_family="MERVL"
exon    complement(16623..16714)
                /note="GRAIL, score = 72.000%, comment = good"
                /evidence=not_experimental
        16661..16730
                /note="MZEF, score = 93.5%"
                /evidence=not_experimental
repeat_region   complement(17540..17675)
                /evidence=not_experimental
                /rpt_family="L1R16C"
repeat_region   complement(17805..17864)
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   complement(18076..18098)
                /evidence=not_experimental
                /rpt_family="MLT1C"
exon    18076..18098
                /note="GRAIL, score = 45.000%, comment = marginal shadow"
                /evidence=not_experimental
exon    complement(18219..18349)
                /note="GenScan, score = 2.99%, comment = Internal_exon 131
                bp frame: 1 phase: 2"
                /evidence=not_experimental
repeat_region   complement(18717..19211)
                /evidence=not_experimental
                /rpt_family="MLT1C"
repeat_region   complement(19252..19999)
                /evidence=not_experimental
                /rpt_family="MLT1C"
exon    complement(19269..19350)
                /note="GRAIL, score = 71.000%, comment = good"
                /evidence=not_experimental
exon    complement(19379..19399)
                /note="GRAIL, score = 81.000%, comment = excellent shadow"
                /evidence=not_experimental
exon    complement(19501..19981)
                /note="MZEF, score = 50.6%"
                /evidence=not_experimental
exon    complement(19945..20020)
                /note="Xpound exon prediction, score = 60% (0%)"
                /evidence=not_experimental
repeat_region   20023..20643
                /note="homology = 100.00%, score = 38, counts = 2"
                /evidence=not_experimental
                /rpt_unit=ttcttgcacacttctctgtgttgaggtcggttatcgc
                /rpt_type=tandem
repeat_region   20645..20910
                /evidence=not_experimental
                /rpt_family="MLT1E"
repeat_region   20961..21007
                /evidence=not_experimental
                /rpt_family="MLT1F"
repeat_region   21183..21781
                /evidence=not_experimental
                /rpt_family="MLT1F"
exon    21504..21756
                /note="MZEF, score = 79.4%"
                /evidence=not_experimental
repeat_region   complement(22325..22806)
                /evidence=not_experimental
                /rpt_family="MLT1C"
```

Query Match   9.6%; Score 119; DB 9; Length 170121;
Best Local Similarity 100.0%; Pred. No. 1.8e-17;
Matches 119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db  120352  GCTTCACCTATGCCAGTACAAACACCACTGACAGCTTCATCTCGGAGTTGATCATCCATG 120293
            |||||||||| | |||| ||| ||||| |||| |||||| |||| ||| ||||||||||
Qy     351  ATGTGGAGCCCAGTGAATTCGGGGAACATCAGAGGCAGCGTCGCCTGCCATG 410

Db  120292  ATGTGCAGACCCCAGTGACTCGGGATCTGCGTCAAATGCAGCCTTGCAGAACAGCACCATGGGTTTG 120233
            |||||| ||||| |||||||| |||| |||||||||| |||| ||||||||||||||||||| |||
Qy     411  GATCTGCTTACCTTACCGTCCAAGTT 436
            |||||||||| ||| ||||||||||||
Db  120232  GATCTGCCTTCCTCTCAGTCAAGGT 120207

RESULT 14
LOCUS        AF045450          40205 bp     DNA      linear    PRI  20-MAR-1998
DEFINITION   Homo sapiens chromosome 21q22.3 cosmid Q11M15, complete sequence.
ACCESSION    AF045450
VERSION      AF045450.1   GI:2895783
KEYWORDS     HTG.
SOURCE       Homo sapiens.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 40205)
  AUTHORS    Taudien,S. and Rosenthal,A.
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 40205)
  AUTHORS    Taudien,S., Nordsiek,G., Dagand,E., Hildmann,T., Drescher,B.,
             Weber,J., Rosenthal,A. and Yaspo,M.L.
  TITLE      Direct Submission
  JOURNAL    Submitted (29-JAN-1998) Genome Analysis, Institut for Molecular
             Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1..40205
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="cosmid Q11M15"
     repeat_region   293..586
                     /evidence=not_experimental
                     /rpt_family="AluSc"
     repeat_region   complement(643..964)
                     /evidence=not_experimental
                     /rpt_family="MERVL"
     repeat_region   complement(1053..1176)
                     /evidence=not_experimental
                     /rpt_family="HERV16"
     exon            1818..2889
                     /note="Xpound exon prediction, score = 82% (0%)"
                     /evidence=not_experimental
     repeat_region   complement(3221..3658)
                     /evidence=not_experimental
                     /rpt_family="MLT1C"
     exon            3359..3395
                     /note="Xpound exon prediction, score = 84% (0%)"
                     /evidence=not_experimental
     repeat_region   complement(3874..4176)
                     /evidence=not_experimental
                     /rpt_family="AluJb"
     repeat_region   complement(4330..4433)
                     /evidence=not_experimental
                     /rpt_family="L2"
     exon            4585..4638
                     /note="GRAIL, score = 95.000%, comment = excellent shadow"
                     /evidence=not_experimental
     repeat_region   complement(5113..7899)
                     /rpt_family="L1PA14"
                     /evidence=not_experimental
     exon            complement(6119..6598)
                     /note="GRAIL, score = 63.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   complement(7927..8045)
                     /rpt_family="FLAM_A"
                     /evidence=not_experimental
     repeat_region   8048..8178
                     /rpt_family="L2"
                     /evidence=not_experimental
     exon            8771..8876
                     /note="MZEF, score = 67.4%"
                     /evidence=not_experimental
     exon            complement(8809..9187)
                     /note="GRAIL, score = 55.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   complement(8837..8876)
                     /rpt_family="MER1A"
                     /evidence=not_experimental
     exon            complement(9113..9213)
                     /note="GRAIL, score = 55.000%, comment = good shadow"
                     /evidence=not_experimental
     repeat_region   complement(9190..9258)
                     /rpt_family="L1MC3"
                     /evidence=not_experimental
     repeat_region   9259..9647
                     /rpt_family="L1MC3"
                     /evidence=not_experimental
     exon            complement(10138..10427)
                     /note="Xpound exon prediction, score = 89% (0%)"
                     /evidence=not_experimental
     repeat_region   10662..10875
                     /rpt_family="MLT2E"
                     /evidence=not_experimental
     exon            complement(11072..11156)
                     /note="MER58A"
                     /evidence=not_experimental
     exon            11570..11704
                     /note="GRAIL, score = 86.000%, comment = excellent"
                     /evidence=not_experimental
     repeat_region   11478..11597
                     /rpt_family="L2"
                     /evidence=not_experimental
     exon            complement(11570..11704)
                     /note="MZEF, score = 66.1%"
                     /evidence=not_experimental
     exon            11570..11704
                     /note="GRAIL, score = 99.000%, comment = excellent shadow"
                     /evidence=not_experimental
     repeat_region   complement(11687..11735)
                     /rpt_family="MIR"
                     /evidence=not_experimental
     repeat_region   complement(11774..12320)
                     /rpt_family="MLT1B"
                     /evidence=not_experimental
     exon            complement(11828..11949)
                     /note="GRAIL, score = 50.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   12623..12843
                     /rpt_family="AluSx"
                     /evidence=not_experimental
     repeat_region   13102..13457
                     /rpt_family="THE1B"
                     /evidence=not_experimental
     exon            complement(13481..13519)
                     /note="GRAIL, score = 65.000%, comment = good"
                     /evidence=not_experimental
     repeat_region   15065..15188
                     /rpt_family="HERVL"
                     /evidence=not_experimental
     misc_feature    15130..15378
```

```
--------    Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
--------    Project Information
    Center project name: L20921
    Center clone name: 147_E_11
--------    Summary Statistics
    Sequencing vector: Plasmid; n/a; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 169879 bases at least Q40
    Consensus quality: 172188 bases at least Q30
    Consensus quality: 173133 bases at least Q20
    Insert size: 176000; agarose-fp
    Insert size: 173761; sum-of-contigs
    Quality coverage: 8.3 in Q20 bases; agarose-fp
    Quality coverage: 8.4 in Q20 bases; sum-of-contigs
--------  .
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*           1    28835: contig of 28835 bp in length
*       28836    28935: gap of     100 bp
*       28936    29562: contig of   627 bp in length
*       29563    29662: gap of     100 bp
*       29663    30328: contig of   666 bp in length
*       30329    30428: gap of     100 bp
*       30429    32005: contig of  1577 bp in length
*       32006    32105: gap of     100 bp
*       32106    33128: contig of  1023 bp in length
*       33129    33228: gap of     100 bp
*       33229    34512: contig of  1284 bp in length
*       34513    34612: gap of     100 bp
*       34613    36079: contig of  1467 bp in length
*       36080    36179: gap of     100 bp
*       36180    37483: contig of  1304 bp in length
*       37484    37583: gap of     100 bp
*       37584    39166: contig of  1583 bp in length
*       39167    39266: gap of     100 bp
*       39267    41065: contig of  1799 bp in length
*       41066    41165: gap of     100 bp
*       41166    43203: contig of  2038 bp in length
*       43204    43303: gap of     100 bp
*       43304    47489: contig of  4186 bp in length
*       47490    47589: gap of     100 bp
*       47590    50604: contig of  3015 bp in length
*       50605    50704: gap of     100 bp
*       50705    56370: contig of  5666 bp in length
*       56371    56470: gap of     100 bp
*       56471    64079: contig of  7609 bp in length
*       64080    64179: gap of     100 bp
*       64180    72619: contig of  8440 bp in length
*       72620    72719: gap of     100 bp
*       72720    81566: contig of  8847 bp in length
*       81567    81666: gap of     100 bp
*       81667    96774: contig of 15108 bp in length
*       96775    96874: gap of     100 bp
*       96875   119647: contig of 22773 bp in length
*      119648   119747: gap of     100 bp
*      119748   147772: contig of 28025 bp in length
*      147773   147872: gap of     100 bp
*      147873   174557: contig of 26685 bp in length
*      174558   174657: gap of     100 bp
*      174658   175861: contig of  1204 bp in length.

FEATURES             Location/Qualifiers
     source          1..175861
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-147E11"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1..28835
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    28936..29562
                     /note="assembly_fragment"
     misc_feature    29663..30328
                     /note="assembly_fragment"
     misc_feature    30429..32005
                     /note="assembly_fragment"
     misc_feature    32106..33128
                     /note="assembly_fragment"
     misc_feature    33229..34512
                     /note="assembly_fragment"
     misc_feature    34613..36079
                     /note="assembly_fragment"
     misc_feature    36180..37483
                     /note="assembly_fragment"
     misc_feature    37584..39166
                     /note="assembly_fragment"
     misc_feature    39267..41065
                     /note="assembly_fragment"
     misc_feature    41166..43203
                     /note="assembly_fragment"
     misc_feature    43304..47489
                     /note="assembly_fragment"
     misc_feature    47590..50604
                     /note="assembly_fragment"
     misc_feature    50705..56370
                     /note="assembly_fragment"
     misc_feature    56471..64079
                     /note="assembly_fragment"
     misc_feature    64180..72619
                     /note="assembly_fragment"
     misc_feature    72720..81566
                     /note="assembly_fragment"
     misc_feature    81667..96774
                     /note="assembly_fragment"
     misc_feature    96875..119647
                     /note="assembly_fragment"
     misc_feature    119748..147772
                     /note="assembly_fragment"
     misc_feature    147873..174557
                     /note="assembly_fragment"
     misc_feature    174658..175861
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    48903 a  39540 c  39029 g  46286 t  2103 others
ORIGIN

Query Match     14.2%;  Score 176.4;  DB 2;  Length 175861;
Best Local Similarity   73.3%;  Pred. No. 4.2e-31;
Matches  239;  Conservative   0;  Mismatches  86;  Indels   1;  Gaps   1;

Qy    112 TGTAGGTTCTGGGTCTTGGGTAAATGCAGACCGT-CTCCCAGGCTGGAAGCTCATGTGGG 171
          ||| ||||| |||| |||| || ||| ||||||| |||||||||| || ||||| |||||
Db 120532 TCTAGGTTCTCCGGATCCAGTTATCAGATCATATAGAAGGTCCTCAGAATGTNAACAGTCCTAAA 120473

Qy    172 GGGCTCCCAGGCTCGCCTTCAAACTCGCACCGT-CTCCCAGGGCTGGAAGCTCATGTGGG 230
          || ||| |||||| || |||||||| ||||||| ||||||| || ||||||||||| ||
Db 120472 GGACTCAGAGGCTCACTTCAAACTCGCACCCGTGGGCTCACGGCTGGAAGCTTCTCAGTGGA 120413

Qy    231 CTCTCAGTGACATGTGTGTTGCTAAGCGTCAGGCCCCATGGAGCCCATCCAACATGACC 290
          || ||| |||||| |||| ||||||||| | | ||||||| || | || || |||||||
Db 120412 CTCTTAACCAAAATGGTGGTCCTGAGTCTCACCACCAAGGACCCATCATCACCAACAAC 120353

Qy    291 GCTTCACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATCATCCACA 350
```

COMMENT   Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M, Lu, X., Hulyk, S.W., Hale, S.M.,
            Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
            Richards, S., Gibbs, R.A.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 10 Row: m Column: 23
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis.

FEATURES   Location/Qualifiers
     source   1..1556
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /map="FVB/N"
                     /clone="MGC:7960 IMAGE:3584645"
                     /tissue_type="Mammary tumor. Metallothionien-TGF alpha
                     model. 10 month old virgin mouse. Taken by biopsy."
                     /clone_lib="NCI_CGAP_Mam1"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     CDS   55..867
                     /codon_start=1
                     /product="RIKEN cDNA 2010003D20 gene"
                     /protein_id="AAH04806.1"
                     /db_xref="GI:11435933"
                     /db_xref="LocusID:72058"
                     /translation="MEGSWRDVLAVLVLIAQLTASGSSYQIIEGPQNVTVLKDSEAHF
                     NCIVTHGWKLLMWTLNQMVVLSLITQGPIITNNRFTYASYNSTDSFISELIIHDVQPS
                     DSSVQCSLQNSHGFGSAFLSVQDSIGEGPALPTWAIILLAVAFSLLLIIVLLII
                     FCCCASRREKEESTYQNEIRKSANMRTNKADPETKLKSGKENYGYSSDEAKAAQTAS
                     LPPKSAEVSLPEKRSSSLPYQELNKHQPGPATHPRVSPDIASPQKVRNVTLV"

BASE COUNT   447 a   364 c   367 g   378 t
ORIGIN

     Query Match   15.0%;   Score 185.6;   DB 10;   Length 1556;
     Best Local Similarity   73.8%;   Pred. No. 2.2e-33;
     Matches   236; Conservative   0; Mismatches   84; Indels   0; Gaps   0;

Qy   115 AGGTTCTGGGTCTCGGTAATGAAGTCATAGAAGGCCCCCAGAATGCAACAGTCCTGAAGG 174
            ||| ||| || |||||||||| || |||| ||| |||| ||| |||||| |||||||||
Db   111 AGCTTCCGGATCCAGTTATCAGATCATAGAAGGTCCTCAGAAGTAACAGTCCTAAAGGA 170

Qy   175 CTCCCAGGCGTCGCTTCAACTGCACGTCGTCTCCAGGGCTGGGAAGCTCATCATGTGGGCTCT 234
            ||| |||||| || ||||||| || || || |||||| || ||||| ||| || |||||||
Db   171 CTCAGAGGCTCACTTCAACTGCACCGTGTCTGGGAAGCTTCTCATGTTGGGACTCT 230

Qy   235 CAGTGACATGGTGGTGCTAAGCGTCAGGCCCATGGAGCCCATCATCACCAATGACCGCTT 294
            |||| ||||| |||||||| |||||||||||| ||||||||||| || ||||||||||||
Db   231 TAACCAAATGGTGGTGCTGAGTCTCACCCAAGGACCCATCATCACACAACCGCTT 290

Qy   295 CACCTCTCAGAGGTACACCAGGGCGGGGAACTTCACCTCGGAGATGATCATCCACAAATGT 354
            ||||| |||| |||| |||||| || ||||||||||||| |||||||| |||| || |||
Db   291 CACCTATGCCAGTTACAACAGCACTGACAGCTTCATCTCGGAGTTGATCATCCATGAGGT 350

Qy   355 GGAGCCCAGTGATTCGGGGAACATCAGATGACAGCCTCCAGAACAGTCGTGCCTGCATGGATC 414
            ||| |||||| || ||||| || || || || |||| || || || || ||||| |||
Db   351 GCAGCCCAGTGACTGCGGGGATCCGTGCAATGCAGCCATGGGTTTGGGATC 410

Qy   415 TGCTTACCTTACCGTCCAAG 434
            ||| | ||| || ||| ||||
Db   411 TGCCTCCTTCTCCAGTGCAAG 430

RESULT 10
LOCUS   AX380408   754 bp   DNA   linear   PAT 18-MAR-2002
DEFINITION   Sequence 13 from Patent WO0200710.
ACCESSION   AX380408
VERSION   AX380408.1 GI:19575338
KEYWORDS   .
SOURCE   house mouse.
  ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE   B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 13 03-JAN-2002;
               Amgen, Inc. (US)
FEATURES   Location/Qualifiers
     source   1..754
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS   53..724
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD28222.1"
                     /db_xref="GI:19575339"
                     /translation="MEGSWRDVLAVLVILAQLTASGSSYQIIEGPQNVTVLKDSEAHF
                     NCTVTHGWKLLMWTLNQMVVLSITTQGPIITNNRFTYASYNSTDSFISELIIHDVQPS
                     DSGSVQCSLQNSHGFGSAPLSVQESTYQNEIRKSANMRTNKADPETKLKSGKENYGYS
                     SDEAKAAQTASLPPKSAEVSLPEKRSSSLPYQELNKHQPGPATHPRVSFDIASPQKVR
                     NVTLV"
BASE COUNT   220 a   191 c   175 g   168 t
ORIGIN

Query Match   15.0%;   Score 185.6; DB 6;   Length 754;
Best Local Similarity   73.8%;   Pred. No. 2.2e-33;
Matches 236; Conservative   0; Mismatches   84; Indels   0; Gaps   0;

Qy   115 AGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGCCCCCAGAAATGCAACAGTCTCTGAAGGG 174
          ||||||||||||||||||||||||||| || |||| |||| ||||||||||||||||||||
Db   109 AGCTTCCGGATCCAGTTATCAGATATAGAAGGTCCTCAGAATGTAACAGTCCTCTAAAGGA 168

Qy   175 CTCCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGGTGGAAGCTCATCATGTGGGCTCT 234
          ||||||||||||||||||||||||||||||||| || |||| |||| |||||| ||| ||
Db   169 CTCAGAGGCTCACTTCAACTGCACCGTGACTCACGGCTGGAAGCTTCTCATGTGGACTCT 228

Qy   235 CAGTGACATGGTGGTTGCTCAGGCGTCACAGGCCCATGGAGGCCCACCAAATGACCGGCTT 294
          | ||||||||| | ||| |||| ||||||||| |||||||| |||| ||||| |||||||
Db   229 TAACCAAATGGTTGGTTCCTGAGTCTCACCACCAAGGACCCATCATCACCAACACCGGCTT 288

Qy   295 CACCCTCTCAGAGGTACGACCAGGGGGGACTTCACCTGGGATGATGATCCACAAATGT 354
          | ||||||||||||||| |||| ||| ||||||| || ||| | |||||| ||||
Db   289 CACCTATGCCAGTTACAACACACACACTGACAGCTTCATCTCGGAGTTGATCATCACAATGT 348

Qy   355 GGAGCCCAGTGATTCGGGGAACATCAGATGGCAGCCTTCCAGAACAGTCGCCTGCATGGATC 414
          |||||| ||||| ||| |||| | || ||| |||| || |||||| |||| ||| |||||
Db   349 GCAGCCCAGTGACTCGTGGGATCCGTGCAAATGCAGAACAGCCCATGGGTTTGGATC 408

Qy   415 TGCTTACCTTACCGTCCAAG 434
          | |||||| ||| || |||
Db   409 TGCCTTCCTCTCTCAGTGCCAAG 428

RESULT 11
LOCUS   AX380406   895 bp   DNA   linear   PAT 18-MAR-2002
DEFINITION   Sequence 11 from Patent WO0200710.
ACCESSION   AX380406
VERSION   AX380406.1 GI:19575336
KEYWORDS   .
SOURCE   house mouse.
  ORGANISM   Mus musculus

REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE   B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 11 03-JAN-2002;
               Amgen, Inc. (US)
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
FEATURES   Location/Qualifiers
     source   1..895
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS   53..865
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD28221.1"
                     /db_xref="GI:19575337"
                     /translation="MEGSWRDVLAVLVILAQLTASGSSYQIIEGPQNVTVLKDSEAHF
                     NCTVTHGWKLLMWTLNQMVVLSITTQGPIITNNRFTYASYNSTDSFISELIIHDVQPS
                     DSGSVQCSLQNSHGFGSAPLSVQDSIGEEGPALPTWAIILLAVAFSLLLIIIVLIII
                     FCCCASPREKEESTYQNEIRKSANMRTNKADPETKLKSGKENYGYSSDEARAAQTAS
                     LPPKSAEVSLPEKRSSSLPYQELNKHQPGPATHPRVSFDIASPQKVRNVTLV"
BASE COUNT   250 a   223 c   212 g   210 t
ORIGIN

Query Match   15.0%;   Score 185.6; DB 6;   Length 895;
Best Local Similarity   73.8%;   Pred. No. 2.2e-33;
Matches 236; Conservative   0; Mismatches   84; Indels   0; Gaps   0;

Qy   115 AGGTTCTGGGTCTGGTAATGAAGTCATAGAAGGCCCCCAGAAATGCAACAGTCTCTGAAGGG 174
          ||||||||||||||||||||||||||| || |||| |||| ||||||||||||||||||||
Db   109 AGCTTCCGGATCCAGTTATCAGATATAGAAGGTCCTCAGAATGTAACAGTCCTCTAAAGGA 168

Qy   175 CTCCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGGTGGAAGCTCATCATGTGGGCTCT 234
          ||||||||||||||||||||||||||||||||| || |||| |||| |||||| ||| ||
Db   169 CTCAGAGGCTCACTTCAACTGCACCGTGACTCACGGCTGGAAGCTTCTCATGTGGACTCT 228

Qy   235 CAGTGACATGGTGGTTGCTCAGGCGTCACAGGCCCATGGAGGCCCACCAAATGACCGGCTT 294
          | ||||||||| | ||| |||| ||||||||| |||||||| |||| ||||| |||||||
Db   229 TAACCAAATGGTTGGTTCCTGAGTCTCACCACCAAGGACCCATCATCACCAACACCGGCTT 288

Qy   295 CACCCTCTCAGAGGTACGACCAGGGGGGACTTCACCTGGGATGATGATCCACAAATGT 354
          | ||||||||||||||| |||| ||| ||||||| || ||| | |||||| ||||
Db   289 CACCTATGCCAGTTACAACACACACACTGACAGCTTCATCTCGGAGTTGATCATCACAATGT 348

Qy   355 GGAGCCCAGTGATTCGGGGAACATCAGATGGCAGCCTTCCAGAACAGTCGCCTGCATGGATC 414
          |||||| ||||| ||| |||| | || ||| |||| || |||||| |||| ||| |||||
Db   349 GCAGCCCAGTGACTCGTGGGATCCGTGCAAATGCAGAACAGCCCATGGGTTTGGATC 408

Qy   415 TGCTTACCTTACCGTCCAAG 434
          | |||||| ||| || |||
Db   409 TGCCTTCCTCTCTCAGTGCCAAG 428

RESULT 12
LOCUS   BC004806   1556 bp   mRNA   linear   ROD 07-AUG-2002
DEFINITION   Mus musculus, RIKEN cDNA 2010003D20 gene, clone MGC:7960
               IMAGE:3584645, mRNA, complete cds.
ACCESSION   BC004806
VERSION   BC004806.1 GI:13435932
KEYWORDS   MGC.
SOURCE   house mouse.
  ORGANISM   Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
REFERENCE   1 (bases 1 to 1556)
  AUTHORS   Strausberg,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (21-MAR-2001) National Institutes of Health, Mammalian
               Gene Collection (MGC), Cancer Genomics Office, National Cancer
               Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
               USA
  REMARK   NIH-MGC Project URL: http://mgc.nci.nih.gov

```
LOCUS       AC020851              196900 bp    DNA     linear   HTG 15-JUL-2000
DEFINITION  Mus musculus clone RP21-467L12, WORKING DRAFT SEQUENCE, 33
            unordered pieces.
ACCESSION   AC020851
VERSION     AC020851.2  GI:9212212
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 196900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196900)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jul 15, 2000 this sequence version replaced gi:6686457.
            -----Genome Center
            Center: Joint Genome Institute
            Center code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 1437240
            Center clone name: RPCI-21_467L12
            -----
            Summary Statistics
            Consensus quality: 164618 bases at least Q40
            Consensus quality: 180887 bases at least Q30
            Consensus quality: 183777 bases at least Q20
            Estimated Insert size: 147000; pulse field gel estimation
            Estimated insert size: 193700; sum-of-contigs estimation
            Quality coverage: 5.67 in Q20 bases; pulse field gel estimation
            Quality coverage: 4.3 in Q20 bases; sum-of-contigs estimation.
            NOTE: This is a 'working draft' sequence. It currently
            consists of 33 contigs. The true order of the pieces
            is not known and their order in this sequence record is
            arbitrary. Gaps between the contigs are represented as
            runs of N, but the exact sizes of the gaps are unknown.
            This record will be updated with the finished sequence
            as soon as it is available and the accession number will
            be preserved.
            *        1        1541: contig of 1541 bp in length
            *     1542        1641: gap of unknown length
            *     1642        3492: contig of 1851 bp in length
            *     3493        3552: gap of unknown length
            *     3593        4738: contig of 1146 bp in length
            *     4739        4838: gap of unknown length
            *     4839        7248: contig of 2410 bp in length
            *     7249        7348: gap of unknown length
            *     7349        9178: contig of 1830 bp in length
            *     9179        9278: gap of unknown length
            *     9279       11193: contig of 2215 bp in length
            *    11494       11593: gap of unknown length
            *    11594       13607: contig of 2014 bp in length
            *    13608       13707: gap of unknown length
            *    13708       15946: contig of 2239 bp in length
            *    15947       16046: gap of unknown length
            *    16047       18392: contig of 2346 bp in length
            *    18393       18492: gap of unknown length
            *    18493       21231: contig of 2739 bp in length
            *    21232       21331: gap of unknown length
            *    21332       24662: contig of 3331 bp in length
            *    24663       24762: gap of unknown length
            *    24763       27140: contig of 2378 bp in length
            *    27141       27240: gap of unknown length
            *    27241       29950: contig of 2710 bp in length
            *    29951       30050: gap of unknown length
            *    30051       32245: contig of 2195 bp in length
            *    32246       32345: gap of unknown length
            *    32346       35688: contig of 3343 bp in length
            *    35689       35788: gap of unknown length
            *    35789       40387: contig of 4599 bp in length
            *    40388       40487: gap of unknown length
            *    40488       43580: contig of 3093 bp in length
            *    43581       43680: gap of unknown length
            *    43681       46437: contig of 2757 bp in length
            *    46438       46537: gap of unknown length
            *    46538       50184: contig of 3647 bp in length
            *    50185       50284: gap of unknown length
            *    50285       54572: contig of 4288 bp in length
            *    54573       54672: gap of unknown length
            *    54673       60227: contig of 5555 bp in length
            *    60228       60327: gap of unknown length
            *    60328       68024: contig of 7697 bp in length
            *    68025       68124: gap of unknown length
            *    68125       78610: contig of 10486 bp in length
            *    78611       78710: gap of unknown length
            *    78711       88151: contig of 9441 bp in length
            *    88152       88251: gap of unknown length
            *    88252       96357: contig of 8006 bp in length
            *    96358       96457: gap of unknown length
            *    96358      106227: contig of 9870 bp in length
            *   106228      106327: gap of unknown length
            *   106328      115795: contig of 9468 bp in length
            *   115796      115895: gap of unknown length
            *   115896      126363: contig of 10468 bp in length
            *   126364      126463: gap of unknown length
            *   126464      137189: contig of 10726 bp in length
            *   137190      137289: gap of unknown length
            *   137290      146378: contig of 9089 bp in length
            *   146379      146478: gap of unknown length
            *   146479      161686: contig of 15208 bp in length
            *   161687      161786: gap of unknown length
            *   161787      176149: contig of 14363 bp in length
            *   176150      176249: gap of unknown length
            *   176250      196900: contig of 20651 bp in length.
FEATURES             Location/Qualifiers
     source          1..196900
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP21-467L12"
                     /clone_lib="RPCI mouse PAC library 21"
BASE COUNT    51456 a  43823 c  44608 g  53666 t  3347 others
ORIGIN
```

```
Query Match            15.1%;  Score 187.4;  DB 2;  Length 196900;
Best Local Similarity   73.5%;  Pred. No. 1e-33;
Matches  239;  Conservative   0;  Mismatches  86;  Indels   0;  Gaps   0;

Qy    112 TGTAGGTTGTCTGGGTCTGTGGTAATGAAAGTCAATAGAAGGCCCCAGAATAGAACAGTCCTGAA 171
          | |||||||| ||| | |||||| ||||| | |||| |||||||||||||||||| || |||| ||
Db  67442 TCTAGCTTCCGGGATCCAGTTATCAGATCATAGAGAAGGTCTTCAGAATGTNAACAGTCCTAAA 67501

Qy    172 GGGCTCCCAGGCGTCGCCTTCAACTGCACCGTTCTCCCAGGGCTGGAAGCTCATCATGTTGGGC 231
          | ||| | |||||| ||| ||| ||||||||||| |||||||||||||||||||| |||
Db  67502 GGACTCAGAGGGCTCACTTCAACTGCACCGTGACTCAGGGCTGGAAGCTTCTCATGTTGGGAC 67561

Qy    232 TCTCAGTGACATGGTGGTTGGTGCTAAGCGTCAGGCCCCATGGAGCCATCATCACCAAATGACCG 291
          | | | ||||| |||| |||| |||| | || ||||||||||| ||||||||| || |||||
Db  67562 TCTTAACCAAAATGGTGGTTGCTAGTCTACCACCCAGGAGCCCATCATCACCAACACCG 67621

Qy    292 CTTCACCTCTCAGAGGTACGACCCAGGCGCGGGAACTTCACCTGGGAGATGGATGCATCACAA 351
          |||||||||||||||| | ||||||| |||||||||||| ||||| | ||| ||| |||||
Db  67622 CTTCACCTATGCCAGTTACAGACAGCACACTGACAGCTTCATCTCGGAGTTGATCATCCATGA 67681

Qy    352 TGTGGAGCCCAGTGATTCGGGGAAACATCAGATGCCAGCCTCCAGAAACAGTCGCGCTCGCATGG 411
          ||| |||||| ||| |||| ||| | | || | |||||||| || |||||| | ||| || ||
Db  67682 TGTGCAGCCCAGTGACTCGGGGATCCGTGCAAATGCAGCCTGCGAAAGCAGCCATGGGTTGG 67741

Qy    412 ATCTGCTTACCTTACCGTCCAAGTT 436
          ||||| |||||||| |||| ||||||
Db  67742 ATCTGCCTTCCTCTCAGTGCAAGGT 67766
```

```
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="21"
                    /map="21q22.3"
                    /clone="BAC-291B3, 3' partial"
                    /clone_lib="BAC library"
                    /note="Accession No. AF064857"
repeat_region   1088..1566
                    /note="L1M4"
                    /rpt_family="LINE/L1"
                    /rpt_type=DISPERSED
repeat_region   2399..2495
                    /note="(TCTA)n"
                    /rpt_family="Simple_repeat"
                    /rpt_type=TANDEM
repeat_region   complement(2496..2878)
                    /note="THE1C"
                    /rpt_family="LTR/MaLR"
                    /rpt_type=DISPERSED
repeat_region   2879..2931
                    /note="(TCTA)n"
                    /rpt_family="Simple_repeat"
                    /rpt_type=TANDEM
repeat_region   2932..3004
                    /note="(TCCA)n"
                    /rpt_family="Simple_repeat"
                    /rpt_type=TANDEM
repeat_region   3005..3074
                    /note="(TCCA)n"
                    /rpt_family="Simple_repeat"
                    /rpt_type=TANDEM
gene            3473..19157
                    /gene="SH3BGR"
mRNA            join(<3473..3565,12649..12678,15331..15460,18739..19157)
                    /gene="SH3BGR"
CDS             join(3473..3565,12649..12678,15331..15426)
                    /partial
                    /gene="SH3BGR"
                    /note="Accession No. X93498"
                    /codon_start=1
                    /product="21-Glutamic Acid Rich protein 21-GARP"
                    /protein_id="CAB90445.1"
                    /db_xref="GI:7717370"
                    /translation="GSEKAEEGGETEAOKEGSEDVGNLPEAQEKNEEGEGETATEETEE
                    IAMEGAEGEAEEEETEAEGEEPGEDEDS"
exon            3473..3565
                    /gene="SH3BGR"
                    /number=4
repeat_region   11702..11753
                    /note="MIR"
                    /rpt_family="SINE/MIR"
                    /rpt_type=DISPERSED
exon            12649..12678
                    /gene="SH3BGR"
                    /number=5
repeat_region   complement(13115..13425)
                    /note="AluSx"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
repeat_region   14013..14131
                    /note="(CA)n"
                    /rpt_family="Simple_repeat"
                    /rpt_type=TANDEM
repeat_region   complement(13598..13892)
                    /note="AluY"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
exon            ...
                    /gene="SH3BGR"
repeat_region   complement(11315..11556)
                    /note="AluJo"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
repeat_region   complement(11002..11282)
                    /note="AluSp"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
repeat_region   10980..11000
                    /note="(TTTA)n"
                    /rpt_family="Simple_repeat"
                    /rpt_type=TANDEM
repeat_region   complement(10430..10551)
                    /note="L2"
                    /rpt_family="LINE/L2"
                    /rpt_type=DISPERSED
repeat_region   complement(10078..10429)
                    /note="THE1B"
                    /rpt_family="LTR/MaLR"
                    /rpt_type=DISPERSED
repeat_region   complement(8865..8964)
                    /note="AluJo"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
repeat_region   complement(8482..8762)
                    /note="L2"
                    /rpt_family="LINE/L1"
                    /rpt_type=DISPERSED
repeat_region   complement(6887..7067)
                    /note="L1MD3"
                    /rpt_family="LINE/L1"
                    /rpt_type=DISPERSED
repeat_region   5542..5829
                    /note="AluSx"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
repeat_region   9131..9424
                    /note="AluSq"
                    /rpt_family="SINE/Alu"
                    /rpt_type=DISPERSED
repeat_region   complement(9723..9882)
                    /note="L2"
                    /rpt_family="LINE/L2"
                    /rpt_type=DISPERSED
repeat_region   complement(9989..10077)
                    /note="L2"
                    /rpt_family="LINE/L2"
                    /rpt_type=DISPERSED
```

```
Query Match     25.7%;  Score 318.6;  DB 9;  Length 340000;
Best Local Similarity   98.8%;  Pred. No. 7.3e-65;
Matches  321;  Conservative   0;  Mismatches  4;  Indels  0;  Gaps  0;

Qy    115 AGGTTCTGGGTCTGGTGAATGAAGTCATAGAAGGCCCCCAGAAATGCAACAGTCCTGAAGGG 174
          ||||||||||||||||| ||||||||||||||||||||| |||||||||| |||||||||||
Db 268453 AGGTTCTGGGTCTGGTTAATGAAGTCATAGAAGGCCCCCAAAATGCAAGAGTCCTGAAGGG 268512

Qy    175 CTCCCAGGCTCGCTTCAACTGCACGTCTCCCAGGGCTGGAAGCTCATCATGTGGGGCTCT 234
          |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db 268513 CTCCCAGGCTCGCTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGGCTCT 268572

Qy    235 CAGTGACATGGTGGTTCGCTAAGCGTCAGGGCCCATGGAGGCCCATCATTCACCAAATGA 294
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db 268573 CAGTGACATGGTGGTTGCTAAGCGTCAGGGCCCATGGAGGCCCATCATTCACCAAATGA 268632

Qy    295 CACCTCTCAGAGGTACGACCAGGGGGGGGGAACTTCACCTCGGAGATGAGATCACACAAATGT 354
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db 268633 CACCTCTCAGAGGTACGACCAGGGGCGGGGAACTTCACCTCGGAGATGAGATCACACAAATGT 268692

Qy    355 GGAGCCCAGTGATTCGGGGGAACATCAGATGCAGCTCCAGAACAGTCGCCTGCATGGATC 414
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db 268693 GGAGCCCAGTGATTCGGGGGAACATCAGATGCAGCTCCAGAACAGTCGCCTGCATGGATC 268752

Qy    415 TGCTTACCTTACCCGTCCAAGTTATG 439
          |||||||||||||||||||| | ||
Db 268753 TGCTTACCTTACCCGTCCAAGGTGTG 268777
```

```
LOCUS       HS21C080             340000 bp    DNA     linear   PRI 24-MAY-2000
DEFINITION  Homo sapiens chromosome 21 segment HS21C080.
ACCESSION   AL163280 AP001735 BA000005
VERSION     AL163280.2  GI:7717369
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 340000)
  AUTHORS   Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
            Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
            Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
            Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
            Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
            Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
            Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,
            Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordsiek,G.,
            Hornischer,K., Brandt,P., Scharfe,M., Schoen,O., Desario,A.,
            Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,
            Hennig,S., Riesselmann,L., Dagand,E., Wehrmeyer,S., Borzym,K.,
            Gardiner,K., Nizetic,D., Francis,F., Lehrach,H., Reinhardt,R. and
            Yaspo,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2000) The Chromosome 21 Mapping and Sequencing
            Consortium: * RIKEN Genomic Sciences Center, Human Genome Research
            Group * Institute of Molecular Biotechnology, Genome Analysis *
            Keio University School of Medicine, Dept. of Molecular Biology *
            GBF, Dept. of Genome Analysis * Max-Planck Institute for Molecular
            Genetics (addresses see below)
  COMMENT   The Chromosome 21 Mapping and Sequencing Consortium consists of
            * RIKEN Genomic Sciences Center, Human Genome Research Group, *
            Saganihara 228-8555, Japan.
            * e.mail: sakaki@gsc.riken.go.jp
            * URL: http://hgp.gsc.riken.go.jp/
            and
            * Institute of Molecular Biotechnology, Genome Analysis, *
            Beutenbergstrasse 11, D-07745 Jena, Germany,
            * e.mail: genome.imb-jena.de
            * URL: http://genome.imb-jena.de/
            and
            * Keio University School of Medicine, Dept. of Molecular Biology, *
            Tokyo 160-8582, Japan,
            * e.mail: shimizu@dmb-med.keio.ac.jp
            * URL: http://adenine.dmb.med.keio.ac.jp/
            and
            * GBF, Dept. of Genome Analysis,
            * Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
            info.genome@gbf.de
            * URL: http://genome.gbf.de/.
            and
            * Max-Planck Institute for Molecular Genetics,
            Ihnestrasse 73, D-14195 Berlin, Germany,
            * e.mail: info-chr21@molgen.mpg.de
            * URL: http://chr21.rz-berlin.mpg.de/.
FEATURES             Location/Qualifiers
     source          1..340000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="P1160L9, 5' partial"
                     /clone_lib="RPCI1,3-5 PAC library"
                     /note="Accession No. AF121897"
     source          73410..243533
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="P70I24"
                     /clone_lib="RPCI1,3-5 PAC library"
                     /note="Accession No. AF064860"
     source          <1..125946
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="P160L9, 5' partial"
                     /clone_lib="RPCI1,3-5 PAC library"
                     /note="Accession No. AF121897"
     source          224137..>340000
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="P206A10, 3' partial"
                     /clone_lib="RPCI1,3-5 PAC library"
                     /note="Accession No. AF121782"
     source          286628..>340000
```

RESULT 8

Query Match 25.7%; Score 318.6; DB 9; Length 142742;
Best Local Similarity 98.8%; Pred. No. 7e-65;
Matches 321; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy    115 AGGTTCTGGGTCTGGGTANTGAAGTCATAAGAAGGCCCCCAGAANTGCCAACAGTCCTGAAGGG 174
          ||||||||||||||||| |||||||||||| |||||||||||| ||||||||||||||||||||
Db  44273 AGGTTCTGGGTCTGGGTAATGAAGTCATAGAGAAGGCCCCCAAAANTGCCAACAGTCCTGAAGGG 44332

Qy    175 CTCCCAGGCGTCGCGTTCAACTGCACGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCTCT 234
          ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db  44333 CTCCCAGGCGTCGCGTTCAACTGCACCGTCTCCCAGGGCTGGAAGCTCATCATGTGGGCTCT 44392

Qy    235 CAGTGACAAYGGTGGTGGTCAGGCCGTNAAGCCTCAGCCCCATCATTCACCAATGACCGGCTT 294
          |||||||| ||||||| | |||  || ||||||||||||||||||| ||||||||||||| |||
Db  44393 CAGTGACAATGGTGGTGCTAAGCGTCAGGCCCATCATCATCACCAATGACCGCTT 44452

Qy    295 CACCTCTCAGAGGTACGACCAGGGCGGGGGAACTTCACCTCGGAGAGATGATCATCCACAATGT 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  44453 CACCTCTCAGAGGTACGACCAGGGCGGGGGAACTTCACCTCGGAGAGATGATCATCCACAATGT 44512

Qy    355 GGAGCCCAGTGAATTCGGGGAACATCAGATGCAGCCTCCAGAACAGTCGCCTGCATGGAGATC 414
          ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  44513 GGAGCCCAGTGAATTCGGGGAACATCAGATGCAGCCTCCAGAAACAGTCGCCTGCATGGAGATC 44572

Qy    415 TGCTTACCTTACCGTCCAAGTTANG 439
          ||||||||||||||||||||| | ||
Db  44573 TGCTTACCTTACCGTCCAAGGTGTG 44597
```

```
     exon            20000..21135
                     /note="MZEF, score = 51.4%"
                     /evidence=not_experimental
     exon            21150..21275
                     /note="xpound exon prediction, score = 74% (0%)"
                     /evidence=not_experimental
     exon            complement(19228..19361)
                     /note="MZEF, score = 92.8%"
                     /evidence=not_experimental
     repeat_region   complement(18643..24647)
                     /rpt_family="HERV17"
                     /note="GRAIL, score = 66.000%, comment = good"
                     /evidence=not_experimental
     exon            complement(18397..18584)
                     /note="GRAIL, score = 51.000%, comment = good shadow"
                     /evidence=not_experimental
     repeat_region   complement(18335..18632)
                     /rpt_family="LTR17"
                     /evidence=not_experimental
     repeat_region   complement(18004..18334)
                     /rpt_family="AluSg1"
                     /evidence=not_experimental
     repeat_region   complement(17633..17969)
                     /rpt_family="MER34"
                     /evidence=not_experimental
     repeat_region   18352..18420
                     /rpt_family="HERV17"
                     /evidence=not_experimental
```

```
repeat_region   complement(140..415)
                /evidence=not_experimental
                /rpt_family="MLT2E"
repeat_region   complement(668..881)
                /evidence=not_experimental
                /rpt_family="MER58A"
exon            complement(1078..1162)
                /evidence=not_experimental
                /note="GRAIL, score = 86.000%, comment = excellent shadow"
exon            1114..1403
                /evidence=not_experimental
                /rpt_family="MLT1I"
repeat_region   1481..1603
                /evidence=not_experimental
                /rpt_family="MLT1I"
repeat_region   1576..1710
                /evidence=not_experimental
                /rpt_family="L2"
exon            1576..1710
                /evidence=not_experimental
                /note="MZEF, score = 66.1%"
exon            1576..1710
                /evidence=not_experimental
                /note="GRAIL, score = 99.000%, comment = excellent"
repeat_region   complement(1693..1741)
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   complement(1780..2326)
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   complement(1834..1955)
                /evidence=not_experimental
                /rpt_family="MLT1E"
exon            2629..2831
                /evidence=not_experimental
                /note="GRAIL, score = 50.000%, comment = good shadow"
repeat_region   3012..3122
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   3123..3478
                /evidence=not_experimental
                /rpt_family="MLT1J"
repeat_region   3479..3618
                /evidence=not_experimental
                /rpt_family="THE1B"
exon            complement(3502..3540)
                /evidence=not_experimental
                /note="GRAIL, score = 65.000%, comment = good shadow"
exon            4674..4781
                /evidence=not_experimental
                /note="GRAIL, score = 42.000%, comment = marginal"
exon            4712..4749
                /evidence=not_experimental
                /note="Xpound exon prediction, score = 70% (0%)"
repeat_region   5385..6104
                /evidence=not_experimental
                /rpt_family="HERVL"
exon            6682..6751
                /evidence=not_experimental
                /note="MZEF, score = 95.1%"
exon            6746..6751
                /evidence=not_experimental
                /note="GenScan, score = 1.83%, comment = Initial_exon 6 bp,
                frame: 1 phase: 0"
repeat_region   6901..7339
                /evidence=not_experimental
                /rpt_family="MLT1C"
repeat_region   7347..7457
                /evidence=not_experimental
                /rpt_family="LTR16C"
repeat_region   complement(7561..7696)
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   complement(7819..8220)
                /evidence=not_experimental
                /rpt_family="MLT1J"

exon            8097..8119
                /note="GRAIL, score = 45.000%, comment = marginal"
                /evidence=not_experimental
repeat_region   complement(8738..9232)
                /evidence=not_experimental
                /rpt_family="MLT1C"
exon            complement(9273..9420)
                /note="GRAIL, score = 71.000%, comment = good"
                /evidence=not_experimental
exon            9290..9371
                /note="GRAIL, score = 81.000%, comment = excellent shadow"
                /evidence=not_experimental
repeat_region   9966..10041
                /note="homology = 100.00%, score = 38, counts = 2"
                /evidence=not_experimental
                /rpt_type=tandem
                /rpt_unit=tccttgcacacttctcgttgagaggtcggttatcgc
repeat_region   10244..10664
                /evidence=not_experimental
                /rpt_family="MLT1E"
repeat_region   10665..11028
                /evidence=not_experimental
                /rpt_family="MLT1P"
repeat_region   11318..11417
                /evidence=not_experimental
                /rpt_family="L1MC/D"
repeat_region   11503..11818
                /evidence=not_experimental
                /rpt_family="MLT1C"
repeat_region   12902..13180
                /evidence=not_experimental
                /rpt_family="AluSg"
exon            complement(13057..13161)
                /note="MZEF, score = 76.1%"
                /evidence=not_experimental
repeat_region   complement(13226..13364)
                /evidence=not_experimental
                /rpt_family="L1MC2"
repeat_region   complement(13613..13908)
                /evidence=not_experimental
                /rpt_family="AluSx"
repeat_region   complement(14317..14614)
                /evidence=not_experimental
                /rpt_family="LTR37A"
exon            complement(14962..15036)
                /note="MZEF, score = 93%"
                /evidence=not_experimental
repeat_region   15255..15299
                /evidence=not_experimental
                /rpt_family="L1PA6"
exon            complement(15469..15497)
                /note="Xpound exon prediction, score = 628 (0%)"
                /evidence=not_experimental
repeat_region   16045..16768
                /evidence=not_experimental
                /rpt_family="MER21A"
exon            complement(16555..16655)
                /note="GRAIL, score = 54.000%, comment = good"
                /evidence=not_experimental
repeat_region   16823..16850
                /evidence=not_experimental
                /rpt_family="MER34"
repeat_region   complement(16851..17208)
                /evidence=not_experimental
                /rpt_family="MIR"
repeat_region   17209..17632
                /rpt_family="THE1A"
                /evidence=not_experimental
repeat_region   17209..17632
                /rpt_family="MLT1J"
                /evidence=not_experimental
```

```
RESULT 6
LOCUS       AX380404       1195 bp    DNA    linear    PAT 18-MAR-2002
DEFINITION  Sequence 9 from Patent WO0200710.
ACCESSION   AX380404
VERSION     AX380404.1  GI:19575334
KEYWORDS
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE     B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 9 03-JAN-2002;
            Amgen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1195
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     CDS             53..1165
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD28220.1"
                     /db_xref="GI:19575335"
                     /translation="MEGSWRDVLAVLVILIAQLIASGSSYQIIEGPQNVTLVKDSEAHF
                     NCTVTHGWKLLMWTLNQMVVLSLTTQGPIITINRRFTYASYNSTDSFISELIIHDVQPS
                     DSGSVQSLIQNSHGFGSAFLSVQVMGTLNIPSNNLIVIEGEPNVTCVAVGWHATSLPDI
                     SWELEVPVSHSSYNSFLEPGNMRVLSVLDLTPLGNGTLITCVARELKDLQASKSLTVNL
                     TVVQPPPDSIGEEGPALPTWAILLLAVAFSLLLIIVLIIFCCCASRREKEESTY
                     QNEIRKSANMRTNKADPETKLKSGKENYGYSSDEAKAQTLASPLPPKSAEVSLPEKRSS
                     SLPYQELNKHQPGPATHPRVSFPDIASPQKVRNVTLV"
BASE COUNT      313 a    304 c    291 g    287 t
ORIGIN
```

```
Qy    Query Match       30.7%;    Score 380.6;   DB 6;   Length 1195;
      Best Local Similarity    62.4%;    Pred. No. 1.1e-79;
      Matches 702;   Conservative   0;   Mismatches 354;   Indels  69;   Gaps   4;

Qy      115 AGGTTCTGGGTCTCGGTAATGAAGTGCATAGAAGGCCCCCAGAATGCAACAGTTCCTGAAGG 174
            |||  ||| ||||  ||||| |  ||| ||||||||||| || ||||  |||||  |||||
Db      109 AGCTTCCGGGATCCAGTTATCAGATCATAGAGAAGGTTCCTCAGAAATGTNAACAGTCCTTAAAGGA 168

Qy      175 CTCCCAGGCTTCGCTTCACACTGCACCGTCTCCCAGGCGTCGGAAGCTCATCATGTCTGGGGCTCT 234
            | ||  |||||||  ||||||| ||| ||| ||||||||||||||||| |||||||||||||||
Db      169 CTCAGAGGCTCACTTCAACTGCACCGTCACGGCTGGTGAAGCTTCTCATGTGGACTCT 228

Qy      235 CAGTGACATGTGGTGGTGCTAAGCGTCAGGGCCCATGGAGCCCATCATCACCAATGACCGCTT 294
            |||||||  ||  || ||| || || ||||||||||| ||||||||||||||||||||| ||||
Db      235 AAGTGGAAAGGAAAAACTACGGGTACAGTTCAAGCTTGAGGCAAAGGCGCACAGACTGCATC 294

Qy      295 CACCTCTCAGAGGTTACGACCAGGCGGGGGAACTTCACCTGGGAGATGAGATCACCAAATGT 354
            ||||||| |||||  || ||||||||| ||||||| || ||||||||||||||||| ||||
Db      289 CACCTATGCCAGTTACACAGCACTGACAGCTTCATCTCGGAGTTGATCATCCATGATGT 348

Qy      355 GGAGCCCACACTGATTCGGGGAACATCAGATGCAGCAGCTCCAGAACAGTCGCCTGCATGGATC 414
            || |||||||| | ||||  ||||||||||||||||||||||||||| |||||||| ||| ||||
Db      349 GCAGCCCAGTGACCTCGGGGATCCGTGCAAATGCAGACACCATGGGGTTGGATC 408

Qy      415 TGCTTACCTTACCGTCCAAGTTATGGGAGAGCGTGTTCATTCCCAGTGTTAAATCTTGTAGT 474
            || |||||| ||||| || ||| ||||||||||||||||||||||||||||| ||||||||
Db      409 TGCCTTCCTCTCAGTGCAGTCACATGGGACCCTGAACAATTCCTAGCCAACAACCTTATAGT 468

Qy      475 CGCTGAGAATGAACCTTGTGAAGTTACTTGTTCTACCCTCACACCCGGCTCCCGGA 534
            |  ||| ||||||||||  ||||  |||| ||| ||||| |||||| |||||||||
Db      469 CACTGAGGGTGAACCCTGTAAATGTGACTTGGCTGGACCTGGACCTGGTTCCACTCCCGGA 528

Qy      535 TATTTCCTTGGGAGCTTCGGGTCTCCGTGGTCAGCCATTCAACTATTATTTTGTTCCGGAGCC 594
            |||||||||||||||||||||| | |||| |||||||||||| ||| |||| || ||||||||
Db      529 TATTTCCTGGGAGCTTGAGGTTACAATTCCTTTCTCCGGAGCC 588

Qy      595 CAGCGACCTTCAAAGTGCCAGTTGAGCACATCCTGGCTCTGACCCCCACAGAGCAATGGGACTTT 654
            |||  || ||||||||||||| ||||| || ||||||||| | ||||| ||||| ||||||  ||
Db      589 GGGCAACTTTATGAGGGTGCTTGAGGTTCCTGGACCCTCACACCACTGGGCACACTGGGGACCTT 648

Qy      655 GACTTGCGTGGCGGTACCTGGAAGAGCCCTGAAGGCCCGGAAGTGTGCAACTGTAAATCTCAC 714
            |||||| |||  || ||||||||||| || |||||||||| |||| ||| || |||| | ||
Db      649 GACTTGGTGGCCAGAGGTGAAGGACTTGCAGGCCAGCAAGTGTCCTTAACTGTCAAACCTGAC 708

Qy      715 TGTGATTCGGTGTCCCAAAGACACTCGAGGTGGTGTATTAATTATTCCAGGTGTGATTATCAAG 774
            | ||| | || ||||||| |||  | ||||||||||||||||||||||||||||||||||||| |
Db      709 TGTGGTTCAGCCTCCACCTGACATTGGAGTTGGAGGA 744

Qy      775 TTTACCGAGTTTAGGTTTTTCATTGCCTACTTGGGGCAAAGTTGGACTGGACTAGCAGG 834
                      ||||||||||||||||||||||||||||||||||||||||||
Db      745 ------AGGCCCAGCAACTGCGCAGCCATCATCCTGCTGGGACTAGCAGG 792

Qy      835 CACCAGCTTCTGACGGCCGACGGTGACTCTTACAATACGGCTCGTGCTGCTGGGCCGGTCG 894
            |||||||||||| ||||||||||||| ||||| ||| |||| | || ||||||| |||| |
Db      835 CACCCTGAACATTCCTAGCAACAACCTTATAGT 894

Qy      895 TTGTTGTTGGGCTGCAACTGGTCGTCGCCGTTGTTTCTGCGTCTGAAGAGAAAAGAGGATT 954
            |||  ||| |||| |||||| ||| | || |||||||||||| ||  ||| |||| | ||
Db      895 TTGTTGTTGGGCTGCAACTGGTCGTCGCTGGTGTTTCTGCGTCTGAAGAGAAAAGAGGATT 954

Qy      955 TCGTATTCAATTTCAAAAGAAATCTGAAAAGAGAACAGAACAACAAAGACAGATCCGGA 1014
            || ||| |||| ||| ||||  ||  || ||||||||||||||||| ||||  |||| ||
Db      955 AAGTGGAAAGGAAAAACTACGGGTACAGTTCAAGCTTGAGGCAAAGGCGCACAGACTGCATC 1014

Qy     1015 AAGTGGAAATGAAAAACTCGGCGTACAAATTCAGATGAACAAAACCACAGAAACCGCTTC 1074
            || ||||| |  ||||| || | || ||||||||||||| |||||||| |||||||||||
Db      900 CTGCAAACATGAGGCACA------ACAAAGCAGATTCCGGAGGAAGAAGTTAA 945

Qy     1075 TCTCCCTCCCCAAAATCTGTGAATCACGTCCTGCATGAACAAAAGAAAAGAAAAAGAGTAGTGTGCGCC 1134
            |||||||| | | ||||||||| |||||||||||||||||||||||||||| ||||||| | ||
Db     1006 TCTCCCTCCTTAAAATCGGCTGTAGCTCAGAAGTCAAGTCCTTCCAGAAAAAACGGCAGTA------GCCT 1059

Qy     1135 TCCTCACCAGCGGGGCTGATCAACGTCCACCCCAGGCCAGCAAGTCATCCAACAGGCTTCTTT 1194
            |||||| ||| ||||| |||||| ||||| | ||||| ||| |||| || || |||||| |
Db     1060 TCCTTATCAGGAACTCAATAAACATCAGCCCCGGTCCAGCAACTTCATCCCACCGGGGTTTCCTT 1119

Qy     1195 TAATCTGGCCCAGTCCTTCGAGAAGGTCAGTAATAACAACTGTAGTGTA 1239
            || |||||| || || ||| ||||| |||| || || |||||||| ||
Db     1120 TGACATCGCCAGTCCTCAGAAGGTCAGAAATGTGACTTTAGTGTA 1164
```

```
RESULT 7
LOCUS       AF121782      142742 bp    DNA    linear    PRI 02-FEB-1999
DEFINITION  Homo sapiens chromosome 21q22.3 PAC 206A10, complete sequence.
ACCESSION   AF121782
VERSION     AF121782.1  GI:4210991
KEYWORDS    HTG.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 142742)
  AUTHORS   Taudien,S., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
            Schattevoy,R., Weber,J., Schilling,M., Menzel,U., Yaspo,M.L. and
            Rosenthal,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JAN-1999) Genome Analysis, Institute for Molecular
            Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES             Location/Qualifiers
     source          1..142742
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="21"
                     /map="21q22.3"
                     /clone="PAC 206A10"
     repeat_region   3..118
                     /evidence=not_experimental
                     /rpt_family="MIR"
```

RESULT 5
LOCUS AX380402
LOCUS AX380402 1139 bp DNA linear PAT 18-MAR-2002
DEFINITION Sequence 7 from Patent WO0200710.
ACCESSION AX380402
VERSION AX380402.1 GI:19575332
KEYWORDS .
SOURCE   human.
 ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
 AUTHORS Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
 TITLE   B7-like molecules and uses thereof
 JOURNAL Patent: WO 0200710-A 7 03-JAN-2002;
    Amgen, Inc. (US)
FEATURES   Location/Qualifiers
 source   1..1139
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
 CDS   1..1134
    /note="unnamed protein product"
    /codon_start=1
    /protein_id="CAD28219.1"
    /db_xref="GI:19575333"
    /translation="MVAGAMENRDPPGSGSGNEVIEGPQNARVLKGSQARFNCTVSQG
    WKILIMWALSDMVVLSVREMEPIITNDRFTSQRYDQGGNLTSEMIIHNVEPSDSGNIRC
    VSHSSYYVPEPSDLQSAVSILALTPQSNGTLICVATWKSLKARKSATVNLTVIRCPQ
    SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLL
    DTGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLITPTCLTIRCCCRRCCGC
    NCCCRCFCCRRKRGNLKKRRQTKKLROKVEMKIPAIIQMNKRPQTPLLSLNPVNPV
    ILKNETVAVALLITSGLIINVHPGQQVIHRLLLIWPVLRSVIQL"
BASE COUNT 290 a 300 c 283 g 266 t
ORIGIN

Query Match 85.8%; Score 1064; DB 6; Length 1139;
Best Local Similarity 97.2%; Pred. No. 5.4e-242;
Matches 1102; Conservative 0; Mismatches 10; Indels 22; Gaps 1;

Qy 106 AGAAGCTGTAGGTCTGGGTCTGGTGAATGAAGTCATAGAAGGCCCCCAGAATGCAACAGT 165
    ||| | ||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db 106 AGAAG----TAGGTCTGGGTCTGGTGAATGAAGTCATAGAAGGCCCCCAGAATGCAACAGT 165

Qy 1186 GGCTTCTTTTAAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATA 1239
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1185 GGCTTCTTTTAAATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATA 1138

RESULT 4
AX380398

LOCUS   AX380398   1168 bp   DNA   linear   PAT 18-MAR-2002
DEFINITION Sequence 3 from Patent WO0200710.
ACCESSION   AX380398
VERSION   AX380398.1 GI:19575328
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE   B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 3 03-JAN-2002;
            Amgen, Inc. (US)
FEATURES   Location/Qualifiers
     source   1..1168
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS   8..1168
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="CAD28217.1"
                     /db_xref="GI:19575329"
                     /translation="MYAGAMENRDPDPGSGSGSGNEVIEGPQNARVLKGSQARNCTVSQG
                     WKLIMWALSDMVLSVRPMEPIITNDRFTSQRYDQGGNFTSEMIIHNVEPSDSGNIRC
                     SLQNSRLHGSAVLTVQVWMGELFIPSVNLVVAAENEPCEVTICLPSHWTWLPDISWELGLL
                     VSHSSSYFVPEPSDLQSAVSILALTPQSNGTLTCVAIWKSLKARKAHVLIVIRCPQ
                     DTGGGINIPGVLSSLPSLGFSLPTWGKVGLGLAGTMLTIPTCTLTIRCCCRRRCCGC
                     NCCCRCCPCCRKRGRFRIQFQKKSEKEKTNKETETESGNENSGYNSDEQKTTDTASLP
                     PKSCESSDPEQRNSSSCGPPHQRADQRPPRPASHPQASFNLASPEKVSNTTVV"
BASE COUNT   299 a   302 c   288 g   279 t
ORIGIN

Query Match   90.2%;   Score 1119;   DB 6;   Length 1168;
Best Local Similarity   99.1%;   Pred. No. 4.7e-255;
Matches 1125;   Conservative 0;   Mismatches 10;   Indels 0;   Gaps 0;

```
Qy   106 AGAAGCTGTAGGTTCTCGGTCTGGTAATGAAGTCATAGAAGGCCCCAGAAATGCAACAGT 165
             |||| |  ||||| |||||| ||||||||||||||||||||||||  |||||||| |||
Db    34 AGACCCACCGGTTCTGGGTCTGGGTAATGAAGTCATAGAAGGCCCCAAAAATGCAAGAGT 93

Qy   166 CCTGAAGGGGCTCTCCAGGCTCGCGTTCACTCGACCGTCTCCCAGGGCTGGAAGCTCATCAT 225
             ||||||||||| |||||| |||| |||||||| |||||||||||||||||||||| ||||||
Db    94 CCTGAAGGGGCTCCCAGGGTTCGGTTCACTCACTGCACCGTCTCCCAGGGCTCATCAT 153

Qy   226 GTGGGCTCTCAGTGACATGGTGGTGCTRAAGCCGTCAGGCCCAGTCATCACCAA 285
             ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   154 GTGGGCTCTCAGTGACATGGTGGTGCTGAAGCCGTCAGGCCCATCAACATCACCAA 213

Qy   286 TGACCGCTTCACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGAGATGATCAT 345
             |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   214 TGACCGCTTCACCTCTCAGAGGTACGACCAGGGCGGGAACTTCACCTCGGAGATGATCAT 273

Qy   346 CCACAAGTGGGAGCCCCAGTGAATTCGGGGAACATCAGATGCAGCGCCTCCAGAAAGTGCCCT 405
             ||||||| |||| |||||  |||||||||||||||||||||||||||||||||||||||||
Db   274 CCACAATGTGGAGCCCAGTGAATTCGGGGAACATCAGATGCAGCCAGAAGTGCCCT 333

Qy   406 GCATGGAATCGCTTACCTTACCGTCCAAGTTATGGGAGAACTGTTCATTCCCAGTGTTAA 465
             ||||||| |||| |||||||||||||||||||||||||||| |||||||||||||||||
Db   334 GCATGGAATCTGCTTACCTTACCGTCCAAGTTATGGGAGAACTTGTTCATTCCCAGTGTTAA 393

Qy   466 TCTTGTAGTCGCTGAGAATGAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACCCG 525
             |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   394 TCTTGTAGTCGCTGAGAATGAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACCTG 453

Qy   526 GCTCCCCGGATATTTCCTGGGGAGCCTGCTCGTCTCTGGGTCAGGTATTTATTTTTGT 585
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   454 GCTCCCCGGATATTTCCTGGGGAGCTCGGTCTCCTGGTCAGCCATTCAAGCTATTATTTTTGT 513

Qy   586 TCCGGAGCCCAGGGACCTTCAAAGTGCAGTGAGCCATCCTGGCTCTGACCCCACAGAGCAA 645
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   514 TCCGGAGCCCAGCGACCTTCAAAAGTGCAGTGAGCAATCCTGGCTCTGACCCCACAGAGCAA 573

Qy   646 TGGGACTTTGACTTGCGTGGCCGCTCACCTGGAAGACCCGCAAGTCTGCAACTGT 705
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   574 TGGGACTTTGACTTGCCTGGCTACCTGGAAGAGCCCTGAAGGCCCGCAAGTCTGCAACTGT 633

Qy   706 AAATCTCACTGTGATTCGGTGTCCCCAAGACACTGGAAGGTGGTATTAATTCCAGTGT 765
             |||| ||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db   634 AAATCTCACTGTCATTCGGTGTCCCCAAGACACTGGAGGTGGTATTAATTCCAGTGT 693

Qy   766 ATTATCAAGTTTACCGAGTTTAGGTTTTTCATTGCCTACTTGGGGCAAAGTGGGACTTGG 825
```

```
Qy   356 GAGCCCAGTGAATTCGGGGAACATCAGATGCAGCTCCTGCCTGCATGGATCT 415
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   291 GAGCCCAGTGAATTCGGGGAACATCAGATGCAGCTCCTGCCTGCATGGATCT 350

Qy   416 GCTTACCTTACCGTCCAGTGTTGTTCATTCCCAGTGTTAATCTTGTAGTC 475
             |||||||||||||||||||||| |||||||||||||||||||||||||||
Db   351 GCTTACCTTACCGTCCAGTGTTATGGGAGAACTTGTTCATTCCCAGTGTTAATCTTGTAGTC 410

Qy   476 GCTGAGAATGAAACCTTGTGAAGTTACTTGTTCTACCCTCACACTGGACCGG 535
             ||||||||||||| |||||||||||||||||||||||||||||||||||| |
Db   411 GCTGAGAATGAAACTTGTGAAGTTACTTGTTCTACCCTCACACTGGGAT 470

Qy   536 ATTTCCTGGGGAGCTCGGTCTCCTGGTCAGCCATTCAAAGCTATTATTTTTGTTCCGGAGCCC 595
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   471 ATTTCCTGGGGAGCTCGGTCTCCTGGTCAGCCATTCAAGCTATTATTTGTTCCGGAGCCC 530

Qy   596 AGCGACCTTCAAAAGTGCAGTGAGCAATCCTGGCTCTGACCCCACAGAGCAATGGGACTTTG 655
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   531 AGCGACCTTCAAAAGTGCAGTGAGCAATCCTGGCTCTGACCCCACAGAGCAATGGGACTTTG 590

Qy   656 ACTTGCCTGGCTGCTACCTGGAAGAGGCCCTGAAGCTGCAACTGTAAAATCTCACT 715
             ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   591 ACTTGCGTGGCTACCTGGAAGAGCCCTGAAGGCCCGCAACTGTAAATCTCACT 650

Qy   716 GTGATTCGGTGTCCCCAAGACACTGGAAGGTGGTATTAATTCCAGTGTTATTATCAAGT 775
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   651 GTGATTCGGTGTCCCCAAGACACTGGAGGTGGTATTAATTCCAGTGTTATTATCAAGT 710

Qy   776 TTACCGAGTTTAGGTTTTTTCATTGCCTACTTGGGGCAAAGTTTGGACTACCAGGCC 835
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   711 TTACCGAGTTTAGGTTTTTTCATTGCCTACTTGGGGCAAAGTTTGGGACTACTAGGCC 770

Qy   836 ACCAATGCTTCTGACCGCGACGTGTACTCTTTACAATACGGCTGCTGTCGCGCGCCGTCGT 895
             |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db   771 ACCAGCTTCTGACCGCGAGGTGTACTCTTTACAATACGGCTGCTGTCGCGCGCCGTCGT 830

Qy   896 TGTTGTGTGGGCTGCAACTGCTGCTGCCGTGTTGTTTGTTTCTGTGTAGAAGAAAAGAGGATTT 955
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   831 TGTTGTGGGCTGCAACTGCTGCTGCCGTGTTGTTTGTTTCTGTAGAAGAAAAGAGGATTT 890

Qy   956 CGTAATTCAATTTCAAAAGAAATCTGAAAAAGAAGAGAACCACAGAAAACCGCTTCT 1015
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   891 CGTAATTCAATTTCAAAAAGAAATCTGAAAAAGAGAACAAAACTGAGACACAGAA 950

Qy  1016 AGTGGAAATGAAAAACTCCGGCTACAATTCAGTATGAGAAAAGAAATCACAGAAAACCGCTTCT 1075
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   951 AGTGGAAATGAAAAACTCCGGCTACAATTCAGTATGAACAAAAGAACCACAGACACCGCTTCT 1010

Qy  1076 CTCCCTCCAAATCCTGAATCCAGTGATCCTGAACAGTAGCTAGCGTGGCCCCT 1135
             |||||||||||||||||||||||||||||||||||||||||||||| |||||
Db  1011 CTCCCTCCAAATCCTGTGAATCCAGTGATCCTGAACAGTATGGGAGAGCCTGTTTAA 1070

Qy  1136 CCTCACCAGCGGGGCTGATCAAACGTCCACCCAGCCAGCAAGTCCACAGGCTTCTTTT 1195
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1071 CCTCACCAGCGGGGTGATCAAACGTCCACCCAGCAAGTCATCCACAGGCTTCTTTT 1130

Qy  1196 AATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG 1240
             ||||||||||||||||||||||||||||||||||||||||||||
Db  1131 AATCTGGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTATAG 1175
```

Left column

```
FEATURES        Location/Qualifiers
     source     1..2051
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="PLACE6017788"
                /tissue_type="placenta"
                /clone_lib="PLACE6"
                /note="cloning vector: pME18SFL3"
BASE COUNT     580 a    463 c    474 g    534 t
ORIGIN

Query Match    90.88;    Score 1126.4;  DB 9;   Length 2051;
Best Local Similarity  99.5%;  Pred. No. 8.4e-257;
Matches 1130;  Conservative  0;  Mismatches  6;  Indels  0;  Gaps  0;

Qy  105 CAGAAAGCTGTAGGTTCTGGGTCTGGGTAATGAACAGCAGAATGCAACAG 164
        ||  |||       || |||||||||||||||||||||||||||||||||
Db  192 CAGTGGTGGGACGGTTCTGGGTCTGGGTAATGAACAGCAGAATGCAACAG 251

Qy  165 TCCTGAAGGGCTCCCAGGCTCGCTTCAACTGCACCTCCAGGCTGGAAGCTCATCA 224
        ||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db  252 TCCTGAAGGGCTTCCCAGGCTCGCTTCAACTGCACCTCCAGGCTGGAAGCTCATCA 311

Qy  225 TGTGGGCTCTCAGTGACATGGTGGTGCTTAAGCGTCAGGCCCATCATCACCA 284
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  312 TGTGGGGCTCTCAGTGACATGGTGGTGCTTAAGCGTCAGGCCCATCATCACCA 371

Qy  285 ATGACCGCTTCACCTCTCAGAGGTAGGACCAGGGGCGGGGAACTTCACCTCGGAGATGATCA 344
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  372 ATGACCGCTTCACCTCTCAGAGGTAGGACCAGGGGCGGGGAACTTCACCTCGGAGATGATCA 431

Qy  345 TCCACAATGTGGAGCCCAGTGATTCGGGGACAACATCAGATGCAGCCTCCAGAACAGTCGCC 404
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  432 TCCACAATGTGGAGCCCAGTGATTCGGGGACAACATCAGATGCAGCCTCCAGAACAGTCGCC 491

Qy  405 TGCCATGGATCTGCTTACCTTACCGTCCCAAGTTATGGGGAGGAGCTGTTCATTCCCAGTGTTA 464
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  492 TGCATGGATCTGCTTACCTTACCGTCCCAAGTTATGTGGAGGAGCTGTTCATTCCCAGTGTTA 551

Qy  465 ACTTGTAGTCGGCTGAGAATGAACCTTGTGAAGTTACTTGTTGTCTCACACTGGACCC 524
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  552 ACTTGTAGTCGGCTGAGAATGAACCTTGTGAAGTTACTTGTTGTCTCACACTGGACCC 611

Qy  525 GGCTCCCGGATATTTCCTGGGAGCTCGGTCTCCTGGTCCAGCCATTCAAAGTCTATTATTTG 584
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  612 GGCTCCCGGATATTTCCTGGGAGCTCGGTCTCCTGGTCCAGCCATTCAAAGTCTATTATTTG 671

Qy  585 TTCCGGAGCCCAGGCAGCGACCTTCAAAGTGCAGTGAGCATCCTGACCCCACACAGAGCA 644
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  672 TTCCGGAGCCCAGGCAGCGACCTTCAAAGTGCAGTGAGCATCCTGACCCCACACAGAGCA 731

Qy  645 ATGGGACTTTGACTTGCGTGCGTGGCTACCTGGAAGAGCGCTGAAGGCCCCGCAAGTGTCGCAACTG 704
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  732 ATGGGACTTTGACTTGCGTGCGTGGCTACCTGGAAGAGCGCTGAAGGCCCCGCAAGTGTCGCAACTG 791

Qy  705 TAAATCTCACTGTGATTCGGTGTGTCCCAAGACACTGGAGGTGGTATTAATAATTCCAGGTG 764
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  792 TAAATCTCACTGTGATTCGGTGTGTCCCAAGACACTGGAGGTGGTATTAATAATTCCAGGTG 851

Qy  765 TATTATCAAAGTTTACCGAGTTTTAGGTTTTTCATTGCCTAGTTGGCAAAGTTGGACTTG 824
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  852 TATTATCAAAGTTTACCGAGTTTTAGGTTTTTCATTGCCTAGTTGGCAAAGTTGGACTTG 911

Qy  825 GACTAGGCAGGACCATGCTTCTGACGGCCGACGGTGTACTTCTACAATACGTGTCTGCTGCT 884
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  912 GACTAGGCAGGCACCATGCTTCTGACGGCCGACGGTGTACTTCTACAATACGTGTGGCTCTC 971

Qy  885 GCCGGCCGCTCGTTGTGTTGTGGGCTCCAACTGCTCCGCCGGTTGTTGTTGTCTGTAGAAGAA 944
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  972 GCCGGCCGCTCGTTGTGTTGTGGGCTCCAACTGCTCCGCCGGTTGTTGTTGTCTGTAGAAGAA 1031

Qy  945 AAAGAGGATTTCGTATTCAATTTCAAAAAGAAAATCTGAAAAAGAGAGAGAACAAACAAAGAAA 1004
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1032 AAAGAGGAGATTCGTATTCAATTTCAATTTCAAAAAGAAATCTGAAAAAGAGAGAGAAGAGAAGACAAAGAAA 1091
```

Right column

```
RESULT 3
LOCUS       AX380396               1175 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 1 from Patent WO0200710.
ACCESSION   AX380396
VERSION     AX380396.1  GI:19575326
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE     B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 1 03-JAN-2002;
            Amgen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1175
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             27..1175
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="CAD28816.1"
                     /db_xref="GI:19575327"
                     /translation="MGLVIFLHGSGSGNEVIEGPQNAFVLKGSQARFNCTVSQGWKLI
                     MWALSDMVVLSVRPMEPIITNDRFTSQRYDQGQGNFTSEMIIHNVEPSDSGNYRCRSLQN
                     SRLHCSSAYLIVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLLVSHS
                     SYFVPEPSDLQSAVSILALTPQSNGLTLVCATWKSLRARKSATVNLVIRCPDGTGG
                     GINIPGVLSLSPLSGFSLPTWGKVGLGLAGTMLLTPTCTLTIRCCCCRRRCCGCNCCC
                     RCFCCRRKRGFRIQFOKKSEKEKTNKETETESGNENSGYNSDEQKTDTIASLPPKSC
                     ESSDPEQRNSSCGPPHQRADQRPRPASHPQASFNLASPEKVSNTTVV"
BASE COUNT     295 a    307 c    286 g    287 t
ORIGIN

Query Match    90.68;    Score 1123.4;  DB 6;   Length 1175;
Best Local Similarity  99.8%;  Pred. No. 4.3e-256;
Matches 1124;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy 1005 CTGAGACAGAAAGTGGAAATGAAAAACTCCGGCTACAATTCAGAATGAACAAGAACCACAG 1064
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1092 CTGAGACAGAAAGTGGAAATGAAAAACTCCGGCTACAATTCAGAATGAACAAGAACCACAG 1151

Qy 1065 AAACCCGCTTCTCTCCCTCCCAAAATCCTGTGAATCCAGTGATCCTGAACAAGAACAGTA 1124
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1152 AAACCCGCTTCTCTCCCTCCCAAAATCCTGTGAATCCAGTGATCCTGAACAAGAACAGTA 1211

Qy 1125 GCTGTGTGGCCCCTCCTCACCCAGCGGGCTGATCAAACGTCCACCCAGCGACGAGACAAGTCATCCAC 1184
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1212 GCTGTGTGGCCCCTCCTCACCCAGCGGGCTGATCAAACGTCCACCCAGCGACGAGACAAGTCATCCAC 1271

Qy 1185 AGGGCTTCTTTTTTAAATCTGGCCAGTCCTTGAGAAGGTCAGTAATACAACTGTAGTATAG 1240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1272 AGGGCTTCTTTTTTAAATCTGGCCAGTCCTTGAGAAGGTCAGTAATACAACTGTAGTATAG 1327

Db 1032 AAAGAGGATTTCGTATTCAATTTCAAAAAGAAAATCTGAAAAAGAGAGAGAACAAACAAAGAAA 1091
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Bottom left (second alignment)

```
BASE COUNT     295 a    307 c    286 g    287 t
ORIGIN

Query Match    90.68;    Score 1123.4;  DB 6;   Length 1175;
Best Local Similarity  99.8%;  Pred. No. 4.3e-256;
Matches 1124;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy  116 GGTTCTGGGTCTGGGTAATGAACAGCAGAATGCAACAGTCCTGAAGGGC 175
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db   51 GGTTCTGGGTCTGGGTAATGAACAGCAGAATGCAACAGTCCTGAAGGGC 110

Qy  176 TCCCAGGCTCGCTTCAACTGCACCTCCCAGGCTGGAAGCTCATCATGTGGGCTCTC 235
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  111 TCCCAGGCTCGCTTCAACTGCACCTCCCAGGCTGGAAGCTCATCATGTGGGCTCTC 170

Qy  236 AGTGACATGGTGGTGCTTAAGCGTCAGGCCCATCATCACCAATGACCGCTTC 295
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  171 AGTGACATGGTGGTGCTTAAGCGTCAGGCCCATCATCACCAATGACCGCTTC 230

Qy  296 ACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATCATCCACAATGTG 355
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  231 ACCTCTCAGAGGTACGACCAGGGCGGGGAACTTCACCTCGGAGATGATCATCCACAATGTG 290
```

FEATURES   Location/Qualifiers
source   1..1240
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
CDS   80..1240
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAD28218.1"
            /db_xref="GI:19575331"
            /translation="MERHLLTVPEAVGSSGSGNEVIEGPONATVLKGSQARFNCTVSQG
            WKLIMWALSDMVVLSVRPMEPIITNDRFTSQRIDOGGNFTSEMIIHNVEPSDSGNIRC
            SLQNSRLHGSAYLTVQVMGELFIPSVNLVVAENEPCEVTCLPSHWTRLPDISWELGLL
            VSHSSYYFVPEPSDLQSAVSILALTPQSNGTLTGVATWKSLKARKSATVNLTVIRCPQ
            DTGGGINIPGVLSLPSLGFSLPTWGKVGLGLAGTMLLIPTCILTIRCCCRRRCCGC
            NCCRCCRRKRGFRIOFPQKKSEKEKITNKETETESGNENSGYTNSDEQKTIETASLP
            PKSCESSSDPEORNSSCGPHQRADQORPRPASHPQASFNLASPEKVSNTTVV"
BASE COUNT 318 a   319 c   305 g   298 t
ORIGIN

Query Match   100.0%; Score 1240; DB 6; Length 1240;
Best Local Similarity 100.0%; Pred. No. 8.7e-284;
Matches 1240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 AGGTGTGAGTCAGTCCAGCCAACAGTGTGGATCAGTTTCCTAGGCTGCCATAACAAAGCACAACAT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AGGTGTGAGTCAGTCCAGCCAACAGTGTGGATCAGTTTCCTAGGCTGCCATAACAAAGCACAACAT 60

Qy     61 AACCTGGTGGCTTAGAACAATGGAAAAGGCATTTGCTACGAAGCTGTAGGTTC 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 AACCTGGTGGCTTAGAACAATGGAAAAGGCATTTGCTACGAAGCTGTAGGTTC 120

Qy    121 TGGGTTCTGGTAATGAAGTCATTAGAAGAGGCCCCCCAGAAATGAACAGTCTGAAGGCCTCCA 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TGGGTTCTGGTAATGAAGTCATTAGAAGAGGCCCCCCAGAAATGAACAGTCTGAAGGCCTCCA 180

Qy    181 GGCTCGCTTCAAACTCCACCGTCTCCCAGGCGCTGAGAGCTCATCATGTGGGCTTCAGTGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGCTCGCTTCAAACTCCACCGTCTCCCAGGCGCTGAGAGCTCATCATGTGGGCTTCAGTGA 240

Qy    241 CATGGTGGTGTGCTAAGCGTCAGGCCCATGGAGCCCATGCCATAACAAATGACCGGCTTCACCTC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CATGGTGGTGTGCTAAGCGTCAGGCCCATGGAGCCCATGCCATAACAAATGACCGGCTTCACCTC 300

Qy    301 TCAGAGGTACGACCAGGGGGGGAAACTTCACCTCGGGAGATGAGGCC 360
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    301 TCAGAGGTACGACCAGGGGGGGAAACTTCACCTCGGGAGATGAGGCC 360

Qy    361 CAGTGCATTCGGGGAACATCAGATGCCAGCCTCCAGCAACAGTCGCGCTGCAGGATCTGCTTA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGTGCATTCGGGGAACATCAGATGCCAGCCTCCAGCAACAGTCGCGCTGCAGGATCTGCTTA 420

Qy    421 CCTTACCGTCCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATTCTTGTAGTCGCTGA 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CCTTACCGTCCCAAGTTATGGGAGAGCTGTTCATTCCCAGTGTTAATTCTTGTAGTCGCTGA 480

Qy    481 GAATGAACCTTGTGTAGGTTACTTGTGTCTACCCTCACACTGGACCCGGCTCCCGGATATTTC 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GAATGAACCTTGTGTAGGTTACTTGTGTCTACCCTCACACTGGACCCGGCTCCCGGATATTTC 540

Qy    541 CTGGGAGCTCGGTGTCTCCTGGTCACCCATTCAAGCTATTATTTTTGTTCCGGAGCCCAGCGA 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTGGGAGCTCGGTGTCTCCTGGTCACCCATTCAAGCTATTATTTTTGTTCCGGAGCCCAGCGA 600

Qy    601 CCTTCAAAGTGCAGTGACGCATCCTGGCTCTGACCCCACAGAGCACTTTGACTTG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CCTTCAAAGTGCAGTGACGCATCCTGGCTCTGACCCCACAGAGCACTTTGACTTG 660

Qy    661 CGTGGCTACCTGAAGAGCCTGAAAGGGCCCGCAAGTCTGCAACTGTAAAATCTCACTGTGAT 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 CGTGGCTACCTGAAGAGCCTGAAAGGGCCCGCAAGTCTGCAACTGTAAAATCTCACTGTGAT 720

Qy    721 TCGGTGTCCCAAGACACTGGAGGTGGTTATTAATATTCCAGGTGTGTATTATCAAGTTTACC 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCGGTGTCCCAAGACACTGGAGGTGGTTATTAATATTCCAGGTGTGTATTATCAAGTTTACC 780

Qy    781 GAGTTTAGGTTTTTCATTGCCTACTTGGGGCAAAGTTGGACTTGGACTAGCAGGCCACCAT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAGTTTAGGTTTTTCATTGCCTACTTGGGGCAAAGTTGGACTTGGACTAGCAGGCCACCAT 840

Qy    841 GCTTCTGACCGCGACGTACTTCTTACAATACGGCTTGTCGTGCTGCCGCGGCCCGTCGTTGTTG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GCTTCTGACCGCGACGTACTTCTTACAATACGGCTTGTCGTGCTGCCGCGGCCCGTCGTTGTTG 900

Qy    901 TGGCTGCAACTGCTGCTGCCGACGTGTACTCTTCACAATACCGTGCGTGTTGCTCGTCTGCCGGGTGGATTTCGTTAT 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TGGCTGCAACTGCTGCTGCCGACGTGTACTCTTCACAATACCGTGCGTGTTGCTCGTCTGCCGGGTGGATTTCGTTAT 960

Qy    961 TCAATTTCAAAAGAAAATCTGAAAAAGAGAAGAGAGAAGACAAAACAAAAGACAAGAAACTGAGACAGAAAGTGG 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 TCAATTTCAAAAGAAAATCTGAAAAAGAGAAGAGAGAAGACAAAACAAAAGACAAGAAACTGAGACAGAAAGTGG 1020

Qy   1021 AAATGAAAACTCCGGCTACAATTCAGATGAACAAAAGACCACAGAAACCGGCTTCTCTCCC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AAATGAAAACTCCGGCTACAATTCAGATGAACAAAAGACCACAGAAACCGGCTTCTCTCCC 1080

Qy   1081 TCCCAAAATCCTGTGAATCCAGTGATCCTGAACACAAAGACGTAGCTGTGGCCCTCCTCA 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TCCCAAAATCCTGTGAATCCAGTGATCCTGAACACAAAGACGTAGCTGTGGCCCTCCTCA 1140

Qy   1141 CCACGGGGCTGATCAAGTCCACCCCCAGCGCCCAGAAGTCATCCACAGCGCTTCTTTTAATCT 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||.
Db   1141 CCACGGGGCTGATCAACGTCCACCCCCAGCGCCCAGAAGTCATCCACAGCGCTTCTTTTAATCT 1200

Qy   1201 GGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTAATAG 1240
          |||||||||||||||||||||||||||||||||||||||||
Db   1201 GGCCAGTCCTGAGAAGGTCAGTAATACAACTGTAGTAATAG 1240
```

RESULT 2
AK092516
LOCUS   AK092516   2051 bp   mRNA   linear   PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ35197 fis, clone PLACE6017788, highly similar
            to IGSF5.
ACCESSION   AK092516
VERSION   AK092516.1 GI:21751130
KEYWORDS   oligo capping; fis (full insert sequence).
SOURCE   Homo sapiens placenta cDNA to mRNA, clone_lib:PLACE6
            clone:PLACE6017788.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 2051)
AUTHORS   Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,
            Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y.,
            Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
            Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
            Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K.
            and Isogai,T.
TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 2051)
AUTHORS   Isogai,T. and Yamamoto,J.
TITLE   Direct Submission
JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.

OM nucleic - nucleic search, using sw model

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: April 28, 2003, 23:25:03 ; Search time 3424.45 Seconds
                 (without alignments)
                 10538.186 Million cell updates/sec

Title: US-09-729-264-5
Perfect score: 1240
Sequence: 1 aggtgtgagtccagccaaca.........gtaatacaactgtagtatag 1240

Scoring table: IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1455140287 8 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*

```
1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_man:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	C	Score	Query Match %	Length	DB	ID	Description
1		1240	100.0	1240	6	AX380400	AX380400 Sequence
2		1126.4	90.8	2051	6	AK092516	Homo sapi
3		1123.4	90.8	1175	6	AX380396	Sequence
4		1119	90.2	1168	6	AX380398	Sequence
5		1064	85.8	1139	6	AX380402	Sequence
6		380.6	30.7	1195	6	AX380404	Sequence
7	c	318.6	25.7	142742	9	AF121782	Homo sapi
8	c	318.6	25.7	340000	9	HS21C080	Homo sapi
9	c	187.4	15.1	196900	2	AC020851	Mus muscu
10		185.6	15.0	754	2	AX380408	Sequence
11		185.6	15.0	895	6	AX380406	Sequence
12		185.6	15.0	1556	10	BC004806	Mus muscu
13	c	176.4	14.2	175861	2	AC120346	Mus muscu
14		119	9.6	40205	9	AF045450	Homo sapi
15		119	9.6	170121	9	AF064860	Homo sapi
16		100	8.1	199665	9	AF064857	Homo sapi
17	c	58.4	4.7	143266	2	AC067790	Homo sapi
18	c	58.4	4.7	152276	2	AC112649	Homo sapi
19	c	58.4	4.7	186968	2	AC007917	Homo sapi
20		58.2	4.7	175861	2	AC120346	Mus muscu
21	c	58.2	4.7	196900	2	AC020851	Mus muscu
22	c	56.6	4.6	427	9	HSU63332	Human super
23	c	56.2	4.5	163295	2	AC078808	Homo sapi
24	c	56.2	4.5	205933	9	AC117381	Homo sapi
25	c	55.4	4.5	53908	9	AC120145	Homo sapi
26	c	55	4.4	1061	9	HSCAGCTG	Homo sapi
27	c	55	4.4	123576	9	AC015969	Homo sapi
28		55	4.4	206059	2	AC127383	Homo sapi
29	c	54.2	4.4	138000	9	AC079127	Homo sapi
30		54.2	4.4	157743	2	AC067887	Homo sapi
31	c	54.2	4.4	183000	2	AC025134	Homo sapi
32	c	53.8	4.3	63692	2	AL603750	Human DNA
33	c	53.8	4.3	160541	2	AL359271	Homo sapi
34	c	53.8	4.3	206187	2	AC021172	Homo sapi
35	c	53.4	4.3	73925	9	AP005626	Homo sapi
36	c	53.4	4.3	164239	9	AC016074	Homo sapi
37	c	53	4.3	118052	2	AC118358	Rattus no
38	c	52.8	4.3	138685	2	AC017582	Homo sapi
39	c	52.8	4.3	166249	3	AC093502	Drosophil
40	c	52.8	4.3	185404	3	AC104703	Drosophil
41	c	52.8	4.3	341319	3	AE003536	Drosophil
42	c	52.6	4.2	143501	9	AC104383	Homo sapi
43	c	52.6	4.2	166836	9	AC019317	Homo sapi
44	c	52.4	4.2	44768	9	AL589649	Human DNA
45	c	52.4	4.2	163357	9	AC091559	Homo sapi

ALIGNMENTS

RESULT 1
AX380400
LOCUS   AX380400   1240 bp   DNA   linear   PAT 18-MAR-2002
DEFINITION Sequence 5 from Patent WO0200710.
ACCESSION   AX380400
VERSION   AX380400.1 GI:19575330
KEYWORDS   .
SOURCE   human.
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Welcher,A.A., Sarmiento,U.M., Schultz,H.J. and Chute,H.T.
  TITLE   B7-like molecules and uses thereof
  JOURNAL   Patent: WO 0200710-A 5 03-JAN-2002;

TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
  ADDRESSEE: PENNIE & EDMONDS
  STREET: 1155 Avenue of the Americas
  CITY: New York
  STATE: New York
  COUNTRY: U.S.A.
  ZIP: 10036-2711
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/08/185,432
  FILING DATE: 21-JAN-1994
  CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
  NAME: Misrock, S. Leslie
  REGISTRATION NUMBER: 18,872
  REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (212) 790-9090
  TELEFAX: (212) 869-8864/9741
  TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
  LENGTH: 3771 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: unknown
  MOLECULE TYPE: CDNA
US-08-185-432-23

Query Match   3.3%;   Score 41;   DB 1;   Length 3771;
Best Local Similarity   66.3%;   Pred. No. 0.1;
Matches   59;   Conservative   0;   Mismatches   30;   Indels   0;   Gaps   0;

```
Qy    873 GCTGCTCGCTGCCTGCCCGCCGCGTGCCTGGTTGTTGGCTGCAACTGCCTGCCCGTTTGTTGTTTCT  932
          ||||||||||||||||| | ||||||     ||||| ||||||||||||||||||||| ||||||| ||
Db   1897 GCTGCTCCTGCGCCCTGTTGCTGCTGCTGTTGCATCTGCAGCTGTTGCTGCTGTTGTTGTTGCTGCT 1956

Qy    933 GCTGTAGGAGAAAAAAGAGGATTTCGTTATT  961
          |||||| | |  |  | | |||||| ||||
Db   1957 GCTGTTGCTGTGTCGCGGATGCAGTGGTAAT 1985
```